Implied Volatility Forecasting:

A Comparison of Different Procedures Including Fractionally Integrated Models with Applications to UK Equity Options

Soosung Hwang
Department of Applied Economics, University of Cambridge

and

Stephen E. Satchell
Faculty of Economics and Politics and Trinity College, University of Cambridge

February 1998

Abstract

The purpose of this paper is to consider how to forecast implied volatility for a selection of UK companies with traded options on their stocks. We consider a range of GARCH and log-ARFIMA based models as well as some simple forecasting rules. Overall, we find that a log-ARFIMA model forecasts best over short and long horizons.

Key-words : Implied Volatility, Forecasting, ARFIMA, GARCH, log-ARFIMA, Long Memory, Options, Fractional Integration.

The authors are grateful to Mr. Carl Moss of Dresdner Kleinwort Benson for his suggestions and the insightful comments of two anonymous referees. The first author would like to thank the Newton Trust and Kleinwort Benson for financial support and the second author would like to thank Inquire for financial support.

1. Introduction

The purpose of this paper is to investigate various procedures for forecasting implied volatility. This topic should be of particular interest to the financial community, especially to option traders, and managers of downside-protected funds. There is a vast bibliography in finance on this topic and we refer readers to Day and Lewis (1992), Harvey and Whaley (1992), Engle, Hong, Kane, and Noh (1993), Lamoureux and Lastrapes (1993), Noh, Engle, and Kane (1994), and Hwang and Satchell (1997) for more details on volatility forecasting. In this study, we further the work of Hwang and Satchell (1997) (HS) on long memory volatility processes for forecasting implied volatility.

Fractionally integrated processes, which are a sub-class of long memory processes, have recently attracted considerable attention in volatility studies. Following the introduction of the autoregressive conditional heteroskedasticity (ARCH) model (Engle, 1982) and the popular generalized ARCH (GARCH) model (Bollerslev, 1986), many empirical studies on volatility in finance have reported the extreme degree of persistence of shocks to the conditional variance process. The integrated GARCH (IGARCH) of Engle and Bollerslev (1986) was formulated to capture this effect. However, in the IGARCH model, the unconditional variance does not exist and a shock remains important for the forecasts of the variance for all future horizons. Ding, Granger, and Engle (1992), using the S&P 500 stock market daily closing price index, show that the autocorrelations of the power transformation of the absolute return are quite high for long lags. The autocorrelations may not be explained properly by either an I(0) or an I(1) process. Motivated by these and other findings, Baillie, Bollerslev, and Mikkelsen (1996) proposed the fractionally integrated GARCH (FIGARCH) model by applying the concept of fractional integration to the GARCH model. In the FIGARCH process, the conditional variance decays at a slow hyperbolic rate for the lagged squared innovations. The concept of a fractional process is rather difficult, we shall discuss it later in the text. Its simple property is that whilst the autocorrelations decrease, they decrease very slowly. Thus, the past influences the future in a manner reminiscent of chaotic processes.

Recently, HS investigated model specification and forecasting performance of FIGARCH, log-FIGARCH, autoregressive fractionally integrated moving average (ARFIMA), log-ARFIMA models for both return volatility and implied volatility. They suggested log-ARFIMA models for implied volatility processes. Log-ARFIMA models are well specified and do not need the non-negativity constraints on their parameters. In addition, using out-of-sample forecast tests, HS showed that for the forecasts of implied volatility, log-ARFIMA models using implied volatility are preferred to conventional GARCH models using return volatility.

Leading on from this work, we further investigate log-ARFIMA models for the prediction of implied volatility. For a practical usage of long memory processes in volatility, two modified versions of long memory processes are also suggested: Scaled truncated log-ARFIMA and detrended log-ARFIMA models. For comparative purposes, we use the GARCH(1,1) model and moving average models. In the next section, we describe the data used here and section 3 explains the models used in this study. In section 4, our results follow, and in section 5 we present conclusions. The mathematical properties of the FIGARCH, log-FIGARCH, and scaled truncated log-FIGARCH models seem to be complex, especially for readers who are not familiar with long memory processes. For these readers, we try to avoid excessive mathematical expressions in the main text. However, for interested readers, more detailed explanations are included in the Appendix.

2. Data

For our data, we use two daily variance series; implied variance (IV) and historical return squared (RS). The IV is provided by the London Financial Options and Futures Exchange (LIFFE) and is calculated from the Black and Scholes (1973) option pricing formula. At-the-money option IVs are used and to minimize term structure effects in volatility, options with the shortest maturity but with at least fifteen working days to maturity are used as in Harvey and Whaley (1991, 1992). In this paper IV is used for the log-ARFIMA (0,*d*,1) model. Note that x_t in this study represents the implied standard deviation at time t. Therefore, x_t^2 represents the IV at time t.

The return series of the underlying asset is provided by Datastream. The RS is calculated from the log-return of the underlying asset less the mean log-return. In what follows, we shall use y_t^2 for the RS at time t. More formally, y_t^2 is obtained from log-return series, r_t, as follows:

$$y_t^2 = 250[r_t - \frac{1}{T}\sum_{t=1}^{T} r_t]^2 \ ,$$

where the number 250 is used to annualize the squared return series. This study uses a GARCH(1,1) process to model RS.

The following nine UK equities and their call options data are used: Barclays, British Petroleum, British Steel, British Telecommunication, BTR, General Electric Co, Glaxo Wellcome, Marks and Spenser, Unilever. In addition, American and European call options on FTSE100 are also used. However, in this paper, the results of British Steel and Glaxo Wellcome are the only ones reported[1].

3. Models for Volatility

In this section, we give details of the models used in this study. In addition, estimation methods and other topics related with forecasting will be explained. Two modified log-ARFIMA models are suggested for the forecast of volatility.

3.1 GARCH Models

A GARCH(1,1) model introduced by Bollerslev (1986) for the residual process, y_t, can be expressed as

$$\begin{aligned} y_t &= \xi_t \sqrt{h_t} \\ h_t &= \omega + \beta h_{t-1} + \alpha y_{t-1}^2 \end{aligned} \qquad (3\text{-}1)$$

where $\xi_t \sim N(0,1)$. The interpretation of h_t is that $h_t = E_{t-1}(y_t^2)$. GARCH(1,1) models relate today's volatility, h_t, to yesterday's volatility, h_{t-1}, and the deviations from value yesterday (y_{t-1}^2). It can successfully explain volatility clustering and fat tails in returns, for further discussion see Bollerslev, Chou, and Kroner (1992). It turns out that GARCH models give qualitatively similar results to volatility forecasts as the exponentially weighted moving average (EWMA) models used in the J P Morgan's RiskMetrics. The EWMA model estimates volatility by considering a geometrically declining weighted sum of lagged squared returns.

The log likelihood function of the GARCH(1,1) model is

[1] The results for the other companies can be obtained from authors on request.

(3-2) $$L(\Xi: y_1, y_2, \ldots, y_T) = -0.5T\,ln(2\pi) - 0.5\sum_{t=1}^{T}[ln(h_t) + \frac{y_t^2}{h_t}]$$

where h_t is given by equation (3-1) and $\Xi' = (\omega, \alpha, \beta)$. The likelihood function is maximized using the Berndt, Hall, Hall, and Hausman (1974) algorithm. Weiss (1986) and Bollerslev and Wooldridge (1992) show that even if the assumption that ξ_t is iid $N(0,1)$ is not valid, the quasi maximum likelihood (QML) estimates obtained by maximizing (3-2) are both consistent and asymptotically normally distributed.

The *f*-step-ahead forecast of implied variance from the GARCH(1,1) model is given by

(3-3) $$\begin{aligned} E_t(y_{t+f}^2) &= \omega\sum_{i=0}^{f-1}(\alpha+\beta)^i + (\alpha+\beta)^{f-1}\beta h_t + (\alpha+\beta)^{f-1}\alpha y_t^2 \qquad f > 1 \\ &= \omega\sum_{i=0}^{f-2}(\alpha+\beta)^i + (\alpha+\beta)^{f-1}h_{t+1} \qquad f > 2 \end{aligned}$$

Therefore, when $\alpha+\beta<1$, for large *f*, the conditional expectation of variance can be represented as

(3-4) $$\begin{aligned} E_t(y_{t+f}^2) &= \omega\sum_{i=0}^{f-2}(\alpha+\beta)^i + (\alpha+\beta)^{f-1}h_{t+1} \\ &\approx \omega\sum_{i=0}^{\infty}(\alpha+\beta)^i \qquad \text{as } f \to \infty \\ &= \omega/(1-\alpha-\beta) \end{aligned}$$

Note that $1/(1-\alpha-\beta)$ is always positive for $0<\alpha+\beta<1$. For large forecasting horizons, the forecasts converge to $\omega/(1-\alpha-\beta)$ at an exponential rate. When the unconditional variance is larger than the first step ahead forecast, the forecasts will show a concave form as the forecast horizon increases. On the other hand, when the unconditional variance is smaller than the first step ahead forecast, the forecasts will show a convex form.

3.2 Long Memory (LM) Volatility Models

3.2.1 General Properties of LM Processes

Long memory process was originally developed in the hydrological literature to explain the Hurst phenomenon (Hurst, 1951): river flows into reservoirs have long memory persistence. In the 1960s it was modelled as a discrete time fractional Gaussian noise (FGN) process by Mandelbrot and Van Ness (1968), which itself is a discrete time version of a fractional Brownian motion. In economics, an alternative approach to modelling long-term persistence was proposed in the early 1980s by Granger and Joyeux (1980), as an extension of the more familiar autoregressive integrated moving average (ARIMA) class of processes.[2]

The discrete time LM process (or, fractionally integrated process) is defined as the following discrete time stochastic process

(3-5) $$(1-L)^d x_t = \varepsilon_t$$

where L is the lag operator and ε_t is an iid random variable. By the binomial series expansion,

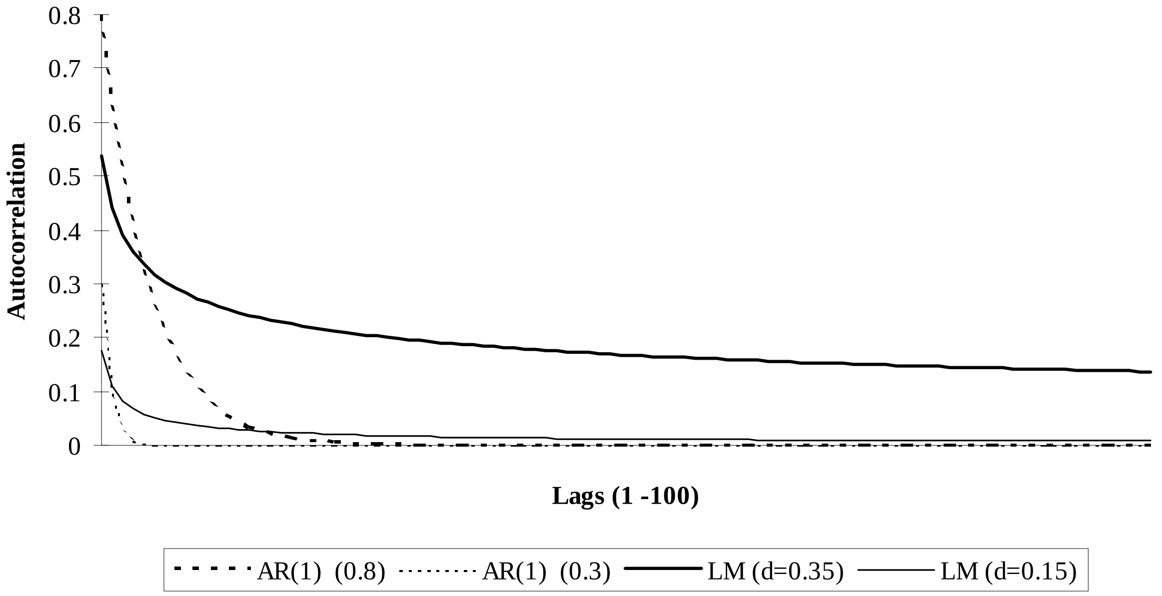
(3-6) $$(1-L)^d = \sum_{j=0}^{\infty}\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}L^j$$

where $\Gamma(.)$ is the gamma function. The general properties of long memory processes are their

[2] The ARFIMA model is generally preferred to the FGN model. The main reason is that the former can describe economic and financial time series better than the latter. Moreover, by capturing both long and short memory, the ARFIMA model is a generalization of the more familiar ARIMA model, and it is easier to use than the FGN model. Furthermore, it need not assume Gaussianity for the innovations.

slowly decaying autocorrelation. Intuitively, this means that an initial shock influences future volatility for a long time. The autocorrelation function of LM processes is proportional to s^{2d-1} as the autocorrelation lag, s, goes to infinity, whilst for standard autoregressive (AR) models the autocorrelation function is $\rho_{AR}(s)=\phi^{s}$, where ϕ is the autoregressive parameter in the AR(1) process. Figure 1 shows that the autocorrelations of LM processes have hyperbolic decay rate, while those of AR(1) processes have exponential decay rate. This is a major difference between conventional short memory processes and LM processes. Appendix 1 shows more mathematical details of LM processes and their properties.

Figure 1 Autocorrelations of AR(1) and Long Memory Processes



Notes: Numbers in the parentheses are the values of the autoregressive parameter in the AR(1) process and the values of the LM parameter in the LM process.

3.2.2 Log-ARFIMA models

Many empirical applications of GARCH models find an apparent persistence of volatility shocks in high frequency financial time series. In order to explain the persistence, Engle and Bollerslev (1986) introduce an integrated GARCH (IGARCH) model. However, it is difficult to ascertain whether or not the apparent persistence indicates integration (see Diebold and Lopez, 1994). Baillie, Bollerslev, and Mikkelsen (1996) suggest the FIGARCH model to capture the long memory present in volatility.

In this study, we use log-ARFIMA models instead of FIGARCH models to model IV. As explained in HS, log-ARFIMA models do not need the non-negativity conditions and their out-of-sample forecasts are not inferior to those of FIGARCH models. The log-ARFIMA (0,*d*,1) model is represented as

(3-7) $\quad (1-L)^{d}\,ln(x_t^2)=\mu+(1+\theta L)\psi_t \qquad 0\le d\le 1$

where ψ_t is a white noise zero mean process ($\psi_t = ln(x_t^2)-E_{t-1}(ln(x_t^2))$). The conditional log-variance of the log-ARFIMA (0,*d*,1) model is

(3-8) $\quad H_t=\mu+\theta\psi_{t-1}+(1-(1-L)^{d})\,ln(x_t^2)$

$$= \mu - \theta H_{t-1} + (1 + \theta L - (1-L)^d) ln(x_t^2)$$
$$= \mu + \theta \psi_{t-1} - \sum_{j=1}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} ln(x_{t-j}^2)$$

where $H_t = E_{t-1}(ln(x_t^2))$. See Appendix 2 for a more detailed explanation of the definition and a discussion of some properties of FIGARCH and log-ARFIMA models.

The quasi maximum likelihood function is

(3-9) $$L(\Xi : x_1, x_2, ..., x_T) = -0.5T\, ln(2\pi) - 0.5 \sum_{t=1}^{T} (ln(\sigma^2) + \frac{(ln(x_t^2) - H_t)^2}{\sigma^2})$$

where H_t is given by equation (3-8) and $\Xi' = (\mu, \mathrm{d}, \theta, \sigma)$. Note that in log-ARFIMA models, we may not assume that innovations are iid normally distributed, and thus, QML estimation is used. This is problematic as there is no reason to believe that our assumption of normality is valid. The Broyden, Fletcher, Goldfarb and Shanno (BFGS) algorithm is used for the maximization of the likelihood function. The *f*-step-ahead conditional log-variance from the log-ARFIMA(0,*d*,1) model at time *t* is given by

(3-10) $$H_{t+f} = \mu - \sum_{j=1}^{f-1} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} H_{t+f-j} - \sum_{j=f}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} ln(x_{t+f-j}^2) \qquad f \geq 2$$

where $H_{t+f} = E_t(ln(x_{t+f}^2))$.

Notice that in log-ARFIMA models, the *f*-step-ahead conditional variance can not be represented as an exponential form of the *f*-step-ahead conditional log-variance; this is a consequence of Jensen’s Inequality. The exponential form of the *f*-step-ahead conditional log-variance is always less than the appropriate forecast. We need a correction factor for the downward biased forecasts; the correction factor is discussed in Appendix 3 .

When fractionally integrated models are estimated, we need pre-sample values and a truncation lag, *m*, of the infinite lag polynomial in log-conditional variances of (3-8). In this study, the unconditional sample log-variance is used for all the pre-sample values as in Baillie, Bollerslev, and Mikkelsen (1996). On the other hand, the truncation lag is set to one hundred as in HS, while previous studies such as Baillie, Bollerslev, and Mikkelsen (1996) and Psaradakis and Sola (1995) set the truncation lag at one thousand for all estimates[3].

We address systematic forecast bias in log-ARFIMA models[4]. An interesting property of long memory volatility processes such as equation (3-7) is that they do not have an unconditional distribution with finite mean (see Appendix 2 for a more detail discussion). In practice, however, we have only finite observations, and a truncation lag, *m*, should be chosen for long memory processes. In this case, the long memory volatility process has an unconditional variance. See Appendix 4 for a more detailed explanation of the existence of an unconditional variance and the systematic forecast bias in a finite sample.

3.2.3 Scaled Truncated Fractionally Integrated Process and Log-ARFIMA Models

In theory, we define long memory volatility models such as FIGARCH or log-ARFIMA models under the assumption that the sample size is infinite. However, in practice, we only have finite samples. As we have seen in the previous subsection, there is a gap between theory and

---

[3] See the explanation of HS. To reduce the calculation time, they used the log-ARFIMA(1,*d*,1) model and searched for the best or at least equivalent truncation lag compared with the 1000 truncation lag. Lags of length, 100, 300, 500, 700, and 1000, were investigated, and the truncation lag which has the maximum log-likelihood value was chosen. The differences in the maximum values between the truncation lags were marginal but the log-ARFIMA(1,*d*,1) model achieved maximum values when the truncation lags were set at 100.

[4] The following explanation applies to all discrete time long memory processes.

actual application, and this issue is focused on whether an unconditional variance exists. The same problem does not arise in conventional ARMA and GARCH models, since the models have short memory and thus the impact of the initial observation becomes negligible even for a small sample size.

Facing these problems in LM models, we suggest scaled truncated long memory models, see Appendix 5. In this model, the sum of the AR coefficients is always forced to one and the unconditional variance does not exist. In addition, the zero mean process is used instead of a drift term, since the assumption of a trend in a volatility process can lead to the non-existence of expected volatility. For forecasting purposes, the standard deviation of the forecasts is expected to be smaller than that of the random walk model, since the forecasts of the scaled truncated log-ARFIMA(0,*d*,1) model are obtained by the weighted average of past variances.

3.2.4 Detrended Log-ARFIMA Models

An alternative and simple method to reduce the systematic forecast bias in log-ARFIMA models is to detrend the forecasts. The detrended *f*-step-ahead conditional log-variance of the log-ARFIMA(0,*d*,1) model can be represented as

$$H_{t+f}^{D} = H_{t+f} - \frac{(H_{t+f^*} - H_{t+1})}{f^*} f \qquad \text{(3-11)}$$

where f^* is the longest forecast horizon, that is, f^*=120 in this study.

This method is based on the stationarity of the volatility process. If there is a downward or upward trend in volatility for a short time period, the detrended method may not be used. If the forecast biases were a linear function of forecast horizons, then detrended log-ARFIMA models would work. However, as we have already noticed in the previous sub-section, the systematic forecast bias changes at a hyperbolic rate over forecasting horizons. Therefore, even if we use this method, there still exists some bias especially in relatively short horizons. Despite all these difficulties, this method has the merit that it is straightforward to use.

3.3 Moving Average Methods

Another frequently used method for the forecast of future implied volatility is the moving average method. We include this procedure as a benchmark. Any sensible forecasting procedure should do just as well as a moving average method. This method is used widely in practice, since traders tend to add a value to the past return volatility to cover their trading costs and other expenses. In this sense, the difference between the implied volatility and return volatility may be called "traders premium".

Using the *n* most recent observations, we can use the following formulae as the forecasts of future volatility.

$$\begin{aligned} FIV_t^{n,RS} &= \sqrt{\frac{1}{n}\sum_{j=0}^{n-1} y_{t-j}^2} \\ FIV_t^{n,IV} &= \sqrt{\frac{1}{n}\sum_{j=0}^{n-1} x_{t-j}^2} \end{aligned} \qquad \text{(3-12)}$$

Note that since the forecasts are not changed for the forecasting horizons in this moving average methods, the statistical properties of the forecasts are the same across all horizons.

4. Out-of-Sample Forecasting Performance Tests

4.1 Forecasting Procedure

Noh, Engle, and Kane (1994) investigate the forecasting performance of the implied and return volatilities in the simulated options. Here, we directly compare the forecasting performance of the alternative models using mean absolute forecast error (MAFE) and mean squared forecast error (MSFE), which are represented as follows.

$$MAFE_f = \frac{1}{240}\sum_{t=1}^{240}\left|FIV_{f,t} - x_{t+f}\right| \quad (4\text{-}1)$$
$$MSFE_f = \frac{1}{240}\sum_{t=1}^{240}\left(FIV_{f,t} - x_{t+f}\right)^2$$

where $MAFE_f$ and $MSFE_f$ represent the MAFE and MSFE at horizon f, respectively, x_{t+f} is the realized implied standard deviation at time $t+f$, and $FIV_{f,t}$ is the forecasted implied standard deviation for horizon f at time t. Note that the $FIV_{f,t}$ for the models used in this study is calculated by equation (3-3) for the GARCH(1,1)-RS model, equation (3-10) for the log-ARFIMA(0,*d*,1)-IV model, equation (3-11) for the detrended log-ARFIMA(0,*d*,1)-IV model, and equation (3-12) for the moving average methods, respectively[5]. In addition, we investigate the forecasting performance of the models over various horizons rather than just one step ahead.

We use a rolling sample of the past volatilities. On day *t*, the conditional volatility of one period ahead, *t*+1, is constructed by using the estimates which are obtained from only the past observations (i.e., 778 observations in this study). By recursive substitution of the conditional volatility, forecasts for up to 120 horizons are constructed. On the next day (*t*+1), using 778 recent observations (i.e., 778 observations from the second observation to the 779 observation), we estimate the parameters again and get another forecast for up to 120 horizons. The estimation and forecasting procedures are performed 240 times using rolling windows of 778 observations[6]. Each forecast is expressed as a standard deviation to be compared with the realized implied standard deviation, and MAFE and MSFE statistics are calculated as in (4-1) above.

4.2 Estimates of GARCH(1,1)-RS, log-ARFIMA(0,*d*,1)-IV, and scaled truncated log-ARFIMA(0,*d*,1)-IV

Table 1 reports the QML estimates of the GARCH(1,1) model using return squared (GARCH(1,1)-RS), the log-ARFIMA(0,*d*,1) model using implied volatility (log-ARFIMA(0,*d*,1)-IV), and the scaled truncated log-ARFIMA(0,*d*,1)-IV for British Steel and Glaxo Wellcome. As frequently found in empirical finance, $\alpha+\beta$ in the GARCH(1,1)-RS model is close to 1 and highly persistent. In this case, long memory processes may be more appropriate for RS than the GARCH(1,1) model.

[5] For the scaled truncated log-ARFIMA(0,*d*,1)-IV model, see equation (A5-6) in Appendix 5.

[6] Our total number of observations are 1148, while the number of observation used is 1138 (778 for estimation and 370 for post sample testing). We first used a 250 out of sample forecast test (in this case, we used all 1148 observations, that is 778 observations for estimation and 380 for post sample testing) for daily estimation. Then, to compare the forecasting performance of daily estimation with those of monthly and quarterly estimation, we used a 240 out of sample forecast test. This is because we need multiples of 20 (for monthly estimation) and 60 (quarterly estimation), see subsection 4.3 and table 3. This paper reports the results of a 240 out of sample forecast test for daily estimation for comparison purpose, since we found no difference between 250 and 240 out of sample forecasting performances.

The middle part of each panel reports the estimates of log-ARFIMA(0,d,1)-IV model. As expected, the drift, μ, is not equivalent to zero and for the truncation lag used in this study, there exists an unconditional log-variance. The lowest parts of panels A and B show the estimates of the scaled truncated log-ARFIMA(0,d,1)-IV model. Note that the long memory parameter of the scaled truncated log-ARFIMA(0,d,1) model, d_{ST}, is smaller than that of the log-ARFIMA(0,d,1) model, d. However, for both models, the estimates of the long memory parameter are significantly different from 0 and 1.

Table 1 Maximum Likelihood Estimates of GARCH(1,1)-RS, Log-ARFIMA(0,d,1)-IV, and Scaled Truncated Log-ARFIMA(0,d,1)-IV models

A. British Steel

Models		ω (μ, δ)	α (θ)	β (d, d_{ST})
GARCH(1,1)-RS	Estimates	0.0004	0.0091	0.9559
	Robust Standard Error	(0.0002)	(0.0079)	(0.0091)
Log-ARFIMA(0,d,1)-IV	Estimates	-0.0509	-0.1309	0.6425
	Robust Standard Error	(0.0135)	(0.0491)	(0.0348)
Scaled Truncated	Estimates	-2.2998	-0.0259	0.5104
Log-ARFIMA(0,d,1)-IV	Robust Standard Error	(0.0295)	(0.0551)	(0.0527)

B. Glaxo Wellcome

Models		ω (μ, δ)	α (θ)	β (d, d_{ST})
GARCH(1,1)-RS	Estimates	0.0019	0.0518	0.9170
	Robust Standard Error	(0.0008)	(0.0128)	(0.0235)
Log-ARFIMA(0,d,1)-IV	Estimates	-0.0221	-0.1792	0.7674
	Robust Standard Error	(0.0072)	(0.0477)	(0.0362)
Scaled Truncated	Estimates	-2.7311	-0.1285	0.7062
Log-ARFIMA(0,d,1)-IV	Robust Standard Error	(0.0295)	(0.0546)	(0.0493)

Notes: Return and implied volatilities from 23 March 1992 to 7 October 1996 for a total of 1148 observations are used.

4.3 Forecasting Performance of the Log-ARFIMA(0,*d*,1)-IV and GARCH(1,1)-RS Models

We summarize our results in table 2. The first two columns in panels A and B of table 2 report the results of an out-of-sample forecasting performance test for British Steel and Glaxo based on the MAFE and the MSFE of the forecasts of implied volatility over the horizons from 1 to 120. Columns 3 and 4 are the results of calculations based on the detrended log-ARFIMA(0,d,1) model of subsection 3.2.4 and the scaled truncated log-ARFIMA(0,d,1) model of subsection 3.2.3 (see Appendix 5 for a more detailed discussion), respectively. The final two columns are obtained from table 5. They describe forecasts based on simple moving averages of RS plus a constant and implied volatility, respectively. The moving average methods were explained in subsection 3.3 and a more detailed explanation on the empirical results are reported in subsection 4.4. We also report in the table the "efficient set" of methods based on smallest MAFE and smallest MSFE. If, for example, $MAFE_{method1}$<$MAFE_{method2}$ and $MSFE_{method1}$>$MSFE_{method2}$, then both methods 1 and 2 are in the efficient set and are reported in bold.

This subsection compares the forecasting performance of the log-ARFIMA(0,d,1)-IV model with that of the GARCH(1,1)-RS model in detail. The first two columns of table 2 show that the MAFE and MSFE of the log-ARFIMA(0,*d*,1)-IV model are smaller than those of the GARCH(1,1)-RS model. In particular, in short horizons, the forecasting performance of the log-ARFIMA(0,*d*,1)-IV model is much better than that of the GARCH(1,1)-RS model. Therefore, for the prediction of implied volatility, the Log-ARFIMA(0,*d*,1)-IV model outperforms the GARCH(1,1)-RS model at least in this context.

Table 2 Forecasting Performance of GARCH(1,1)-RS, Long Memory Volatility Models, and Moving Average Methods

A. British Steel Plc

Fore-casting Hori-zons	GARCH(1,1)-RS		Log-ARFIMA(0,*d*,1) -IV		Detrended Log-ARFIMA(0,*d*,1) -IV		Scaled Truncated Log-ARFIMA(0,*d*,1) -IV		Return Squared (*n*=60) Increased by 0.0661		Implied Volatility (*n*=20)	
	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE
1	0.0308	0.0016	0.0117	0.0003	0.0117	0.0003	**0.0116**	**0.0003**	0.0207	0.0008	0.0142	0.0005
5	0.0288	0.0015	0.0165	0.0006	0.0162	0.0006	0.0156	0.0006	0.0205	0.0008	**0.0153**	**0.0006**
10	0.0270	0.0014	0.0186	0.0007	0.0182	0.0007	0.0174	0.0007	0.0208	0.0008	**0.0165**	**0.0006**
20	0.0255	0.0013	0.0195	0.0007	0.0191	0.0007	**0.0178**	**0.0006**	0.0221	0.0010	0.0181	0.0006
30	0.0259	0.0013	0.0200	0.0008	0.0189	0.0007	**0.0175**	**0.0006**	0.0240	0.0010	0.0188	0.0007
40	0.0274	0.0014	0.0219	0.0008	0.0202	0.0007	0.0190	0.0007	0.0249	0.0010	**0.0176**	**0.0006**
50	0.0256	0.0012	0.0208	0.0007	0.0186	0.0006	**0.0173**	**0.0006**	0.0247	0.0009	0.0182	0.0006
60	0.0250	0.0011	0.0217	0.0007	0.0181	0.0006	**0.0168**	**0.0005**	0.0271	0.0011	0.0200	0.0007
80	0.0262	0.0011	0.0257	0.0010	0.0216	0.0008	**0.0203**	**0.0007**	0.0301	0.0013	0.0210	0.0008
100	0.0255	0.0011	0.0281	0.0012	0.0230	0.0008	0.0229	0.0008	0.0307	0.0013	**0.0227**	**0.0008**
120	0.0260	0.0011	0.0293	0.0012	0.0241	0.0009	**0.0226**	**0.0008**	0.0308	0.0014	0.0230	0.0008

B. Glaxo Wellcome Plc

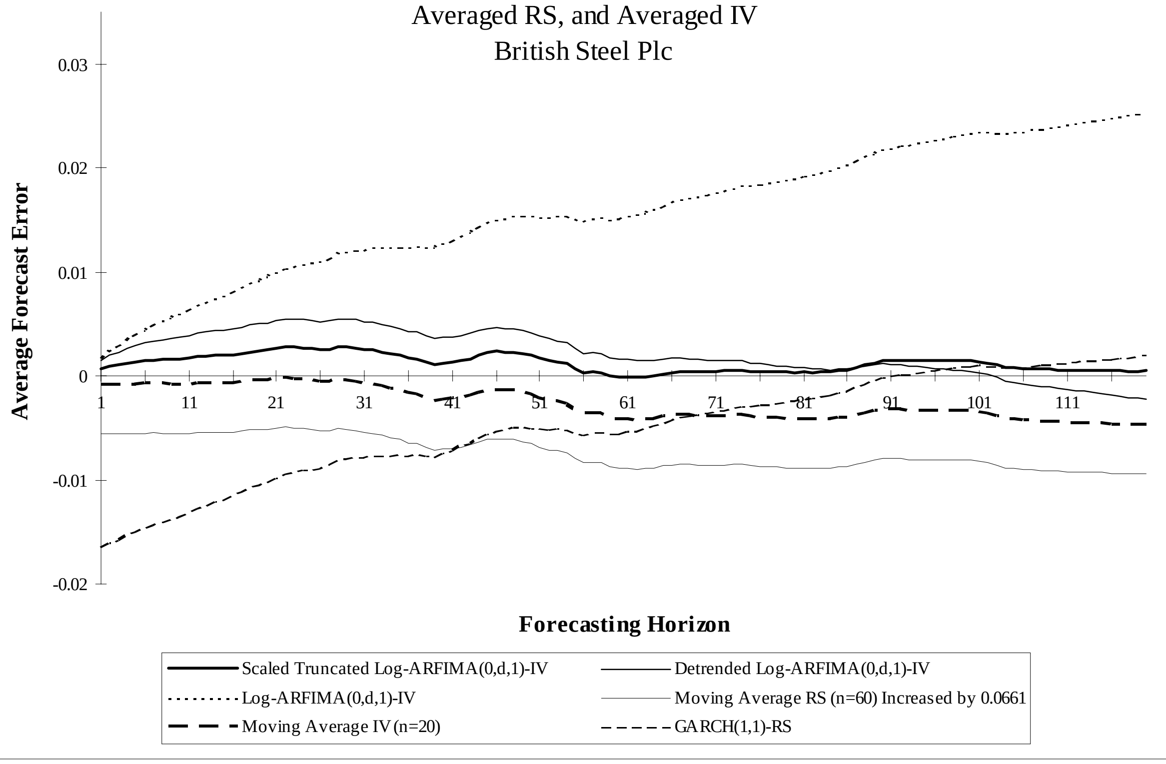
Fore-casting Hori-zons	GARCH(1,1)-RS		Log-ARFIMA(0,*d*,1) -IV		Detrended Log-ARFIMA(0,*d*,1) -IV		Scaled Truncated Log-ARFIMA(0,*d*,1) -IV		Return Squared (*n*=60) Increased by 0.0465		Implied Volatility (*n*=20)	
	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE
1	0.0281	0.0013	**0.0048**	**0.0000**	0.0048	0.0001	**0.0048**	**0.0000**	0.0130	0.0002	0.0075	0.0001
5	0.0351	0.0016	0.0076	0.0001	0.0074	0.0001	**0.0073**	**0.0001**	0.0137	0.0003	0.0089	0.0002
10	0.0416	0.0020	0.0103	0.0002	0.0100	0.0002	**0.0098**	**0.0002**	0.0140	0.0003	0.0101	0.0002
20	0.0457	0.0024	0.0120	0.0002	0.0114	0.0002	**0.0111**	**0.0002**	0.0149	0.0003	0.0114	0.0003
30	0.0490	0.0027	0.0128	0.0003	0.0120	0.0003	**0.0117**	**0.0002**	0.0159	0.0003	0.0120	0.0003
40	0.0538	0.0032	0.0145	0.0003	0.0132	0.0003	0.0129	0.0003	0.0155	0.0003	**0.0123**	**0.0003**
50	0.0564	0.0036	0.0150	0.0004	0.0131	0.0003	0.0128	0.0003	0.0141	0.0003	**0.0119**	**0.0002**
60	0.0574	0.0037	0.0144	0.0003	0.0128	0.0003	0.0125	0.0003	0.0145	0.0003	**0.0121**	**0.0002**
80	0.0587	0.0038	0.0147	0.0004	0.0127	0.0003	**0.0125**	**0.0003**	0.0136	0.0003	**0.0125**	**0.0003**
100	0.0575	0.0037	0.0163	0.0005	0.0140	0.0003	**0.0137**	**0.0003**	0.0162	0.0004	0.0140	0.0003
120	0.0574	0.0036	0.0167	0.0005	0.0152	0.0004	**0.0146**	**0.0004**	0.0170	0.0005	0.0149	0.0004

Notes : GARCH(1,1)-RS forecasts for implied standard deviation (ISD) are obtained using return squared, while Log-ARFIMA(0,*d*,1)-IV forecasts for ISD are calculated using implied volatility. Return and implied volatilities from 23 March 1992 to 7 October 1996 for a total of 1148 observations are used. The most recent 778 observations are used for estimating models and predicting future ISDs over 120 horizons. The results are based on 240 out-of-sample forecasts. Bold numbers represent the smallest MAFE and the smallest MSFE for given forecasting horizons. In the case of a tie or a non-ranking, both are recorded in bold.

The MAFE and MSFE used here show only the magnitude of the forecast error and do not show forecast bias (FB) and forecast standard deviation (FSTD)[7]. Figures 2A and 3A plot the average forecast errors over forecasting horizons for British Steel and Glaxo. During the forecasting period, the realized IV of Glaxo is relatively less volatile than that of British Steel. The magnitude of the average forecast errors tends to increase as forecasting horizons increase for both models. In short horizons, the log-ARFIMA(0,*d*,1)-IV average forecast errors are very small. Over long horizons, the log-ARFIMA(0,*d*,1)-IV forecasts are less biased than the GARCH(1,1)-RS forecasts for Glaxo, while the log-ARFIMA(0,*d*,1)-IV forecasts are more biased than the GARCH(1,1)-RS forecasts for British Steel. This shows that as explained in subsection 3.2.2, a drift term together with the truncation lag may result in a large forecast bias in the log-ARFIMA model.

Figures 2B and 3B plot the FSTDs of the forecasts for the two companies. The log-ARFIMA(0,*d*,1)-IV model has lower FSTD than the GARCH(1,1)-RS model in short forecasting horizons. However, in long horizons, the FSTD of the GARCH(1,1)-RS model is little different from that of the log-ARFIMA(0,*d*,1)-IV model for Glaxo. Although it is not reported in this paper, British Petroleum and Barclays also show that the FSTD of the log-ARFIMA(0,*d*,1)-IV model is lower than that of the GARCH(1,1)-RS model. Our conclusion is that the log-ARFIMA(0,*d*,1)-IV model has less FSTD than the GARCH(1,1)-RS model.

Figure 2A Average Forecast Error of GARCH(1,1)-RS, Log-ARFIMA(0,*d*,1)-IV, Detrended Log-ARFIMA(0,*d*,1)-IV, Scaled Truncated Log-ARFIMA(0,*d*,1)-IV, Averaged RS, and Averaged IV
British Steel Plc



[7] Note that MSFE may be decomposed into the sum of squared forecast bias and forecast variance. For the models such as log-ARFIMA(0,*d*,1)-IV model which have systematic forecast bias, the FB include both the systematic forecast bias and the differences between forecasts and realised IVs for given forecasting horizons. On the other hand, for models such as GARCH(1,1)-RS model which do not have systematic forecast bias, the FB simply represents the sum of the differences between forecasts and realised IVs for a given forecasting horizon.

Figure 2B Standard Deviation of Forecasts of GARCH(1,1)-RS, Log-ARFIMA(0,*d*,1)-IV, Detrended Log-ARFIMA(0,*d*,1)-IV, Scaled Truncated Log-ARFIMA(0,*d*,1)-IV, Average RS, and Average IV
British Steel Plc

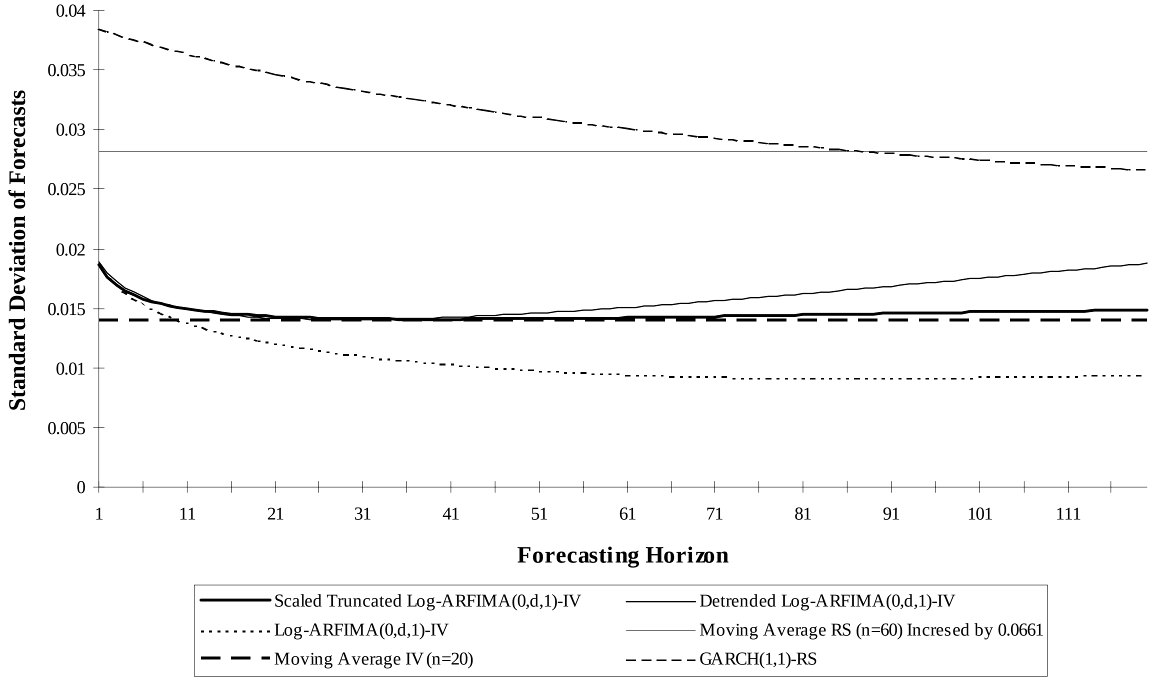
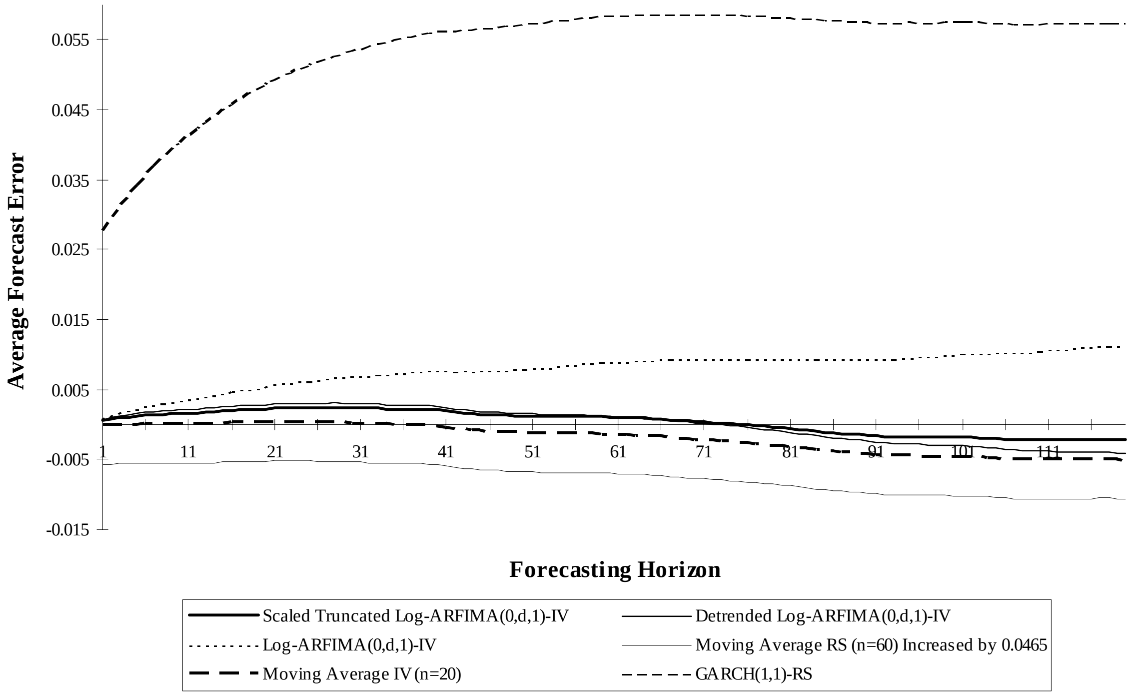
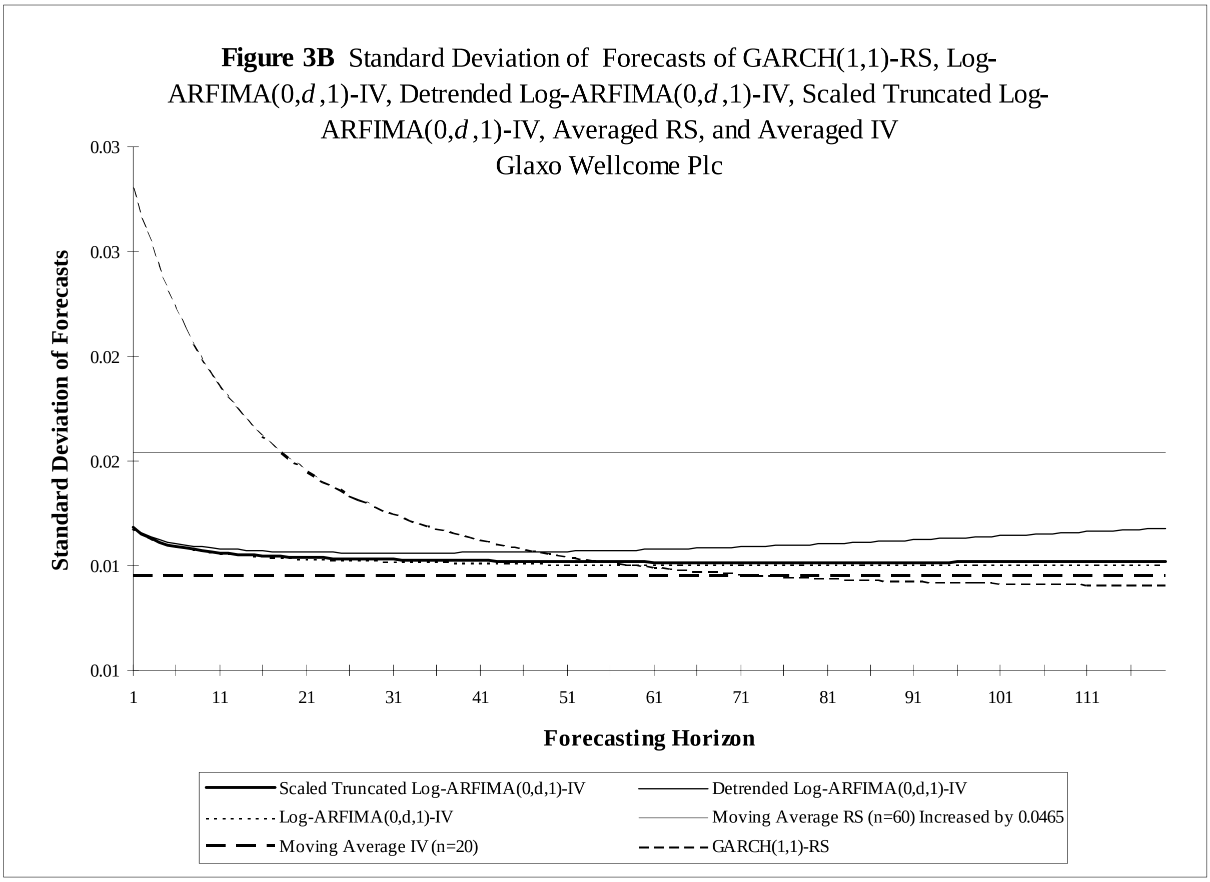


Figure 3A Average Forecast Error of GARCH(1,1)-RS, Log-ARFIMA(0,*d*,1)-IV, Detrended Log-ARFIMA(0,*d*, 1)-IV, Scaled Truncated Log-ARFIMA(0,*d*,1)-IV, Averaged RS, and Averaged IV
Glaxo Wellcome Plc





Notes: Figures 2A and 3A plot average forecast errors over forecasting horizons for British Steel and Glaxo Wellcome and figures 2B and 3B plot forecast standard deviations over forecasting horizons for the two companies. The GARCH(1,1)-RS, log-ARFIMA(0,d,1)-IV, scaled truncated log-ARFIMA(0,d,1)-IV, and detrended log-ARFIMA(0,d,1)-IV models are explained in subsections 3.1, 3.2.2, 3.2.3, and 3.2.4, respectively. The moving average IV (n=20) represents forecasts based on averaged value of last 20 IVs. The moving average RS (n=60) increased by a number (0.0661 for British Steel and 0.0465 for Glaxo Wellcome) represents forecasts based on averaged value of last 60 RSs plus the optimal increase. The moving average methods were explained in subsection 3.3 and a more detailed explanation on the empirical results are reported in subsection 4.4. The MAFE and the MSFE of the forecasts are summarized in table 2.

We need to address the issue of when to re-estimate the models. In practice, daily estimation of a model may be time-consuming work. If there is little difference in forecasting performance between daily estimation and longer estimation intervals, e.g., weekly, monthly, and quarterly, we need not estimate the models daily. For example, for the case of monthly estimation, a model can be estimated once a month and the estimates can be used to forecast implied volatility for the next month.

Table 3 reports the results. For British Steel, the forecasting performance gets better as the estimation interval increases for the GARCH(1,1)-RS model, while it becomes slightly worse for the larger estimation intervals for the log-ARFIMA(0,*d*,1)-IV model. On the other hand, for Glaxo the forecasting performance gets worse as the estimation interval increases for both log-ARFIMA(0,*d*,1)-IV and GARCH(1,1)-RS models. However, we find that the GARCH(1,1)-RS model still does not outperform the log-ARFIMA(0,*d*,1)-IV model and the difference between the forecasting performances from the different estimation intervals is marginal for the log-ARFIMA-IV model. Therefore, on the ground of these results, we can conclude that the log-ARFIMA(0,*d*,1)-IV model need not be estimated daily and can be estimated monthly without particularly increasing the forecasting error.

Table 3 Forecasting Performance of the GARCH(1,1)-RS and Log-ARFIMA(0,*d*,1)-IV Models Considering Estimation Intervals

A. British Steel Plc

Estimation Interval		1 (Daily)		5 (Weekly)		10 (Fortnightly)		20 (Monthly)		60 (Quarterly)	
		MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE
Forecasting Performance of GARCH(1,1)-RS Model	1	0.0308	0.0016	0.0304	0.0015	0.0303	0.0015	0.0294	0.0014	0.0273	0.0013
	5	0.0288	0.0015	0.0285	0.0015	0.0282	0.0015	0.0279	0.0014	0.0261	0.0013
	10	0.0270	0.0014	0.0267	0.0014	0.0266	0.0014	0.0266	0.0014	0.0248	0.0012
	15	0.0255	0.0013	0.0254	0.0013	0.0254	0.0013	0.0250	0.0013	0.0240	0.0011
	20	0.0259	0.0013	0.0257	0.0013	0.0256	0.0012	0.0249	0.0012	0.0240	0.0011
	30	0.0274	0.0014	0.0269	0.0013	0.0265	0.0013	0.0262	0.0012	0.0256	0.0012
	40	0.0256	0.0012	0.0253	0.0012	0.0252	0.0012	0.0252	0.0012	0.0243	0.0011
	60	0.0265	0.0012	0.0265	0.0012	0.0266	0.0011	0.0261	0.0011	0.0237	0.0009
	80	0.0254	0.0010	0.0250	0.0010	0.0250	0.0009	0.0248	0.0009	0.0236	0.0008
	100	0.0255	0.0011	0.0254	0.0010	0.0250	0.0010	0.0245	0.0010	0.0226	0.0008
	120	0.0260	0.0011	0.0258	0.0010	0.0253	0.0010	0.0247	0.0010	0.0242	0.0009
Forecasting Performance of Log-ARFIMA(0,*d*,1) -IV Model	1	0.0117	0.0003	0.0117	0.0003	0.0118	0.0003	0.0117	0.0003	0.0117	0.0003
	5	0.0165	0.0006	0.0165	0.0006	0.0165	0.0006	0.0164	0.0006	0.0165	0.0006
	10	0.0186	0.0007	0.0186	0.0007	0.0186	0.0007	0.0184	0.0007	0.0186	0.0007
	15	0.0195	0.0007	0.0195	0.0007	0.0195	0.0007	0.0195	0.0007	0.0197	0.0007
	20	0.0200	0.0008	0.0200	0.0008	0.0200	0.0008	0.0201	0.0008	0.0204	0.0008
	30	0.0219	0.0008	0.0219	0.0008	0.0219	0.0008	0.0219	0.0008	0.0222	0.0008
	40	0.0208	0.0007	0.0208	0.0007	0.0209	0.0007	0.0211	0.0007	0.0216	0.0008
	60	0.0239	0.0009	0.0242	0.0009	0.0243	0.0009	0.0243	0.0009	0.0247	0.0009
	80	0.0259	0.0010	0.0261	0.0010	0.0264	0.0010	0.0267	0.0010	0.0273	0.0011
	100	0.0281	0.0012	0.0283	0.0012	0.0286	0.0012	0.0293	0.0012	0.0310	0.0013
	120	0.0293	0.0012	0.0295	0.0012	0.0299	0.0012	0.0306	0.0012	0.0320	0.0013

B. Glaxo Wellcome Plc

Estimation Interval		1 (Daily)		5 (Weekly)		10 (Fortnightly)		20 (Monthly)		60 (Quarterly)	
		MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE
Forecasting Performance of GARCH(1,1)-RS Model	1	0.0281	0.0013	0.0281	0.0013	0.0286	0.0013	0.0300	0.0014	0.0319	0.0015
	5	0.0351	0.0016	0.0353	0.0016	0.0360	0.0016	0.0376	0.0018	0.0405	0.0020
	10	0.0416	0.0020	0.0420	0.0020	0.0427	0.0021	0.0443	0.0022	0.0478	0.0026
	15	0.0457	0.0024	0.0462	0.0024	0.0469	0.0025	0.0484	0.0027	0.0521	0.0031
	20	0.0490	0.0027	0.0495	0.0028	0.0501	0.0028	0.0516	0.0030	0.0551	0.0034
	30	0.0538	0.0032	0.0543	0.0033	0.0550	0.0034	0.0561	0.0035	0.0593	0.0039
	40	0.0564	0.0036	0.0569	0.0036	0.0574	0.0037	0.0584	0.0038	0.0613	0.0042
	60	0.0585	0.0038	0.0588	0.0038	0.0593	0.0039	0.0600	0.0040	0.0625	0.0043
	80	0.0583	0.0038	0.0586	0.0038	0.0591	0.0039	0.0596	0.0039	0.0618	0.0042
	100	0.0575	0.0037	0.0578	0.0037	0.0582	0.0037	0.0587	0.0038	0.0608	0.0041
	120	0.0574	0.0036	0.0577	0.0036	0.0581	0.0037	0.0585	0.0037	0.0605	0.0040
Forecasting Performance of Log-ARFIMA(0,*d*,1) -IV Model	1	0.0048	0.0000	0.0048	0.0000	0.0049	0.0000	0.0049	0.0000	0.0049	0.0000
	5	0.0076	0.0001	0.0076	0.0001	0.0076	0.0001	0.0077	0.0001	0.0077	0.0001
	10	0.0103	0.0002	0.0104	0.0002	0.0104	0.0002	0.0104	0.0002	0.0105	0.0002
	15	0.0120	0.0002	0.0120	0.0002	0.0120	0.0002	0.0120	0.0002	0.0122	0.0002
	20	0.0128	0.0003	0.0128	0.0003	0.0128	0.0003	0.0128	0.0003	0.0132	0.0003
	30	0.0145	0.0003	0.0145	0.0003	0.0145	0.0003	0.0146	0.0003	0.0152	0.0004
	40	0.0150	0.0004	0.0150	0.0004	0.0150	0.0004	0.0151	0.0004	0.0160	0.0004
	60	0.0146	0.0004	0.0147	0.0004	0.0146	0.0004	0.0149	0.0004	0.0158	0.0004
	80	0.0147	0.0004	0.0148	0.0004	0.0147	0.0004	0.0149	0.0004	0.0164	0.0005
	100	0.0163	0.0005	0.0164	0.0005	0.0164	0.0005	0.0167	0.0005	0.0185	0.0006
	120	0.0167	0.0005	0.0168	0.0005	0.0167	0.0005	0.0171	0.0005	0.0190	0.0006

Notes : GARCH(1,1)-RS forecasts for implied standard deviation (ISD) are obtained using return squared, while Log-ARFIMA(0,*d*,1)-IV forecasts for ISD are calculated using implied volatility. Return squared and implied volatilities from 23 March 1992 to 7 October 1996 for a total of 1148 observations are used. The most recent 778 observations are used for estimating models and predicting future ISDs over 120 horizons. The results are based on 240 out-of-sample forecasts. For the case of daily estimation, each model is estimated and the forecasts are obtained on the daily basis, whilst for the quarterly estimation, the models are estimated once every 60 days and the estimates are used for the forecasts. Therefore, the number of estimations is 240 for the daily estimation while it is only 4 for the quarterly estimation. Note that forecasting is always performed on a daily basis.

4.4 Forecasting Performance of the Moving Average Methods

We need to investigate the forecasting performance of the practically widely used moving average methods in detail. The moving average forecasts for implied volatility using IV and RS with the *n* most recent observations at time *t*, $FIV_t^{n,IV}$ and $FIV_t^{n,RS}$, are discussed in subsection 3.3. Table 4 reports mean and standard deviation of the forecasts. As expected, for a given moving average lag (*n*), the mean of $FIV_t^{n,RS}$ is smaller than that of $FIV_t^{n,IV}$, while the standard deviation of $FIV_t^{n,RS}$ is larger than that of $FIV_t^{n,IV}$.

Table 4 Mean and Standard Deviation of Moving Average Forecasts of RS and IV

A. British Steel Plc

Moving Average Lag (*n*)		1	5	10	15	20	60
Return Squared	Mean	0.1767	0.1759	0.1753	0.1742	0.1734	0.1711
	STD	0.1443	0.0755	0.0546	0.0479	0.0433	0.0282
Implied Volatility	Mean	0.2428	0.2426	0.2424	0.2422	0.2420	0.2401
	STD	0.0245	0.0198	0.0171	0.0153	0.0140	0.0106

B. Glaxo Wellcome Plc

Moving Average Lag (*n*)		1	5	10	15	20	60
Return Squared	Mean	0.1473	0.1469	0.1464	0.1452	0.1438	0.1415
	STD	0.1346	0.0651	0.0476	0.0397	0.0325	0.0154
Implied Volatility	Mean	0.1938	0.1938	0.1938	0.1937	0.1936	0.1958
	STD	0.0129	0.0118	0.0112	0.0104	0.0095	0.0110

Notes: Return and implied volatilities from 23 March 1992 to 7 October 1996 for a total of 1148 observations are used.

Using the same forecasting procedure in subsection 4.1, we calculate the MSFE and MAFE of the moving average methods. The forecasting performances of $FIV_t^{n,IV}$ and $FIV_t^{n,RS}$ are reported in table 5. Table 5 also reports the forecast performance of $FIV_t^{n,RS}$ increased by some numbers from the original $FIV_t^{n,RS}$. This is because RS is generally less than IV and the original $FIV_t^{n,RS}$ may result in downward FB if unadjusted. The optimal increase (i.e., 0.0661* for British Steel and 0.0465* for Glaxo) is chosen to match mean of $FIV_t^{1,RS}$ with that of $FIV_t^{1,IV}$. FIV_t^{n,RS^*} is used for $FIV_t^{n,RS}$ with this increase. Therefore, FIV_t^{n,RS^*} is the sum of the moving average forecasts at time *t* and the optimal increase which is obtained using all ex-post moving average forecasts. In this sense, FIV_t^{n,RS^*} is not an out-of-sample forecast, but we use it for purposes of comparison. Here, we consciously use ex-post values. This is because the procedure reflects a simple forecasting rule communicated to us by practitioners, using ex-post information allows us to build in practitioner expertise.

Note that the MSFEs of FIV_t^{n,RS^*} with *n*>1 are smaller than those of the original $FIV_t^{1,RS}$. Since MSFE can be decomposed into the sum of squared forecast bias and forecast variance, this can be explained by follows; as the moving average lag (n) increases, the FSTD of $FIV_t^{n,RS}$ reduces, and as the mean of $FIV_t^{n,RS}$ goes to FIV_t^{n,RS^*}, FB decreases. Therefore, the table shows that, when we use moving averaged RS as forecast of future volatility, large n and an appropriate increase should be considered.

FIV_t^{n,RS^*} may have less MAFE and MSFE than the GARCH(1,1)-RS method in the previous section. However, we calculated the optimal increase by "data snooping", and since we do not know how much we increase $FIV_t^{n,RS}$, the simple moving average method may not be preferred to the GARCH(1,1)-RS method. Moreover, even though we choose the optimal increase and a large moving average lag, the forecast performance of the FIV_t^{n,RS^*} does not outperform $FIV_t^{n,IV}$; see the last rows of panels A and B of table 5[8]. Therefore, we can conclude that for the forecast of IV, IV should be used rather than RS.

In addition, we investigate the selection of n for $FIV_t^{n,IV}$. Table 5 shows that for the forecast of short horizons, $FIV_t^{1,IV}$ outperforms $FIV_t^{n,IV}$ with $n>1$. However, for long forecasting horizons, $n=20$ seems to be appropriate[9]. The last rows of panels A and B in table 5 show that MAFE and MSFE tend to decrease as n becomes larger. For large n, there is little difference in MSFE and MAFE and in particular, for Glaxo, some MSFEs and MAFEs of $FIV_t^{60,IV}$ are larger than those of $FIV_t^{n,IV}$ with the smaller n.

Table 5 Forecasting Performance of Moving Average Forecasts of RS and IV

A. British Steel

Average Lag(n)		1		10		20		60	
		MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE
Return Squared	1	0.13383	0.02506	0.07286	0.00752	0.07210	0.00669	0.07170	0.00590
	5	0.13343	0.02518	0.07359	0.00760	0.07178	0.00665	0.07163	0.00588
	10	0.13371	0.02540	0.07401	0.00758	0.07096	0.00662	0.07175	0.00588
	20	0.13206	0.02494	0.07268	0.00730	0.06970	0.00652	0.07123	0.00592
	40	0.13470	0.02551	0.07304	0.00782	0.07127	0.00695	0.07337	0.00626
	60	0.13511	0.02549	0.07481	0.00793	0.07405	0.00723	0.07527	0.00668
	80	0.13497	0.02557	0.07532	0.00833	0.07464	0.00749	0.07585	0.00685
	100	0.13542	0.02551	0.07665	0.00844	0.07615	0.00768	0.07463	0.00679
	120	0.13821	0.02627	0.07863	0.00854	0.07676	0.00778	0.07568	0.00702
Return Squared Increased by 0.02	1	0.12479	0.02282	0.05982	0.00523	0.05513	0.00431	0.05184	0.00343
	5	0.12407	0.02294	0.06051	0.00530	0.05507	0.00427	0.05187	0.00341
	10	0.12415	0.02316	0.06041	0.00528	0.05443	0.00424	0.05195	0.00342
	20	0.12294	0.02271	0.05914	0.00502	0.05249	0.00417	0.05159	0.00347
	40	0.12509	0.02321	0.05973	0.00546	0.05383	0.00452	0.05367	0.00374
	60	0.12522	0.02312	0.06145	0.00550	0.05671	0.00473	0.05631	0.00409
	80	0.12503	0.02320	0.06243	0.00590	0.05789	0.00498	0.05858	0.00425
	100	0.12571	0.02316	0.06410	0.00604	0.06102	0.00520	0.05665	0.00422
	120	0.12830	0.02387	0.06547	0.00608	0.06113	0.00525	0.05725	0.00440
	1	0.11736	0.02137	0.05084	0.00373	0.04275	0.00274	0.03373	0.00176

[8] British Petroleum and Barclays also show that the forecast performance of the FIV_t^{n,RS^*} does not outperform $FIV_t^{n,IV}$.

[9] Although it is not reported in this paper, British Petroleum and Barclays also show that n=20 is an appropriate value.

	5	0.11637	0.02150	0.05125	0.00380	0.04199	0.00270	0.03347	0.00175
Return	10	0.11651	0.02172	0.05069	0.00378	0.04130	0.00266	0.03337	0.00175
Squared	20	0.11613	0.02129	0.04923	0.00354	0.04093	0.00261	0.03388	0.00182
Increased	40	0.11732	0.02171	0.05039	0.00391	0.04125	0.00288	0.03618	0.00201
by 0.04	60	0.11738	0.02155	0.05215	0.00387	0.04401	0.00302	0.04091	0.00229
	80	0.11723	0.02162	0.05304	0.00427	0.04587	0.00327	0.04370	0.00246
	100	0.11793	0.02162	0.05549	0.00444	0.04982	0.00353	0.04309	0.00245
	120	0.12037	0.02227	0.05566	0.00443	0.04876	0.00352	0.04197	0.00258
	1	0.11116	0.02069	0.046156	0.002974	0.03479	0.00188	0.02072	0.00079
	5	0.11036	0.02082	0.045719	0.003055	0.03402	0.00185	0.02046	0.00078
Return	10	0.11033	0.02103	0.045123	0.003029	0.03434	0.00181	0.02077	0.00078
Squared	20	0.11005	0.02063	0.043493	0.002814	0.03523	0.00178	0.02211	0.00087
Increased	40	0.11019	0.02095	0.045859	0.003084	0.03607	0.00195	0.02471	0.00096
by 0.0661*	60	0.11071	0.02069	0.046274	0.002945	0.03485	0.00199	0.02713	0.00114
	80	0.11062	0.02077	0.048408	0.003345	0.03945	0.00225	0.03014	0.00131
	100	0.11118	0.02080	0.050181	0.003556	0.04242	0.00254	0.03071	0.00135
	120	0.11362	0.02139	0.049591	0.00348	0.04034	0.00247	0.03075	0.00140
	1	**0.01238**	**0.00038**	0.01411	0.00050	0.01417	0.00052	0.01449	0.00050
	5	0.01826	0.00084	0.01686	0.00069	**0.01527**	**0.00060**	**0.01534**	**0.00054**
	10	0.02128	0.00104	0.01908	0.00074	0.01646	0.00061	**0.01622**	**0.00056**
Implied	20	0.02310	0.00100	0.01929	0.00069	0.01813	0.00063	**0.01743**	**0.00061**
Volatility	40	0.02252	0.00098	0.01923	0.00064	**0.01756**	**0.00057**	0.01776	0.00060
	60	0.02402	0.00106	0.02192	0.00081	0.01997	0.00070	**0.01883**	**0.00063**
	80	0.02562	0.00108	0.02323	0.00087	0.02099	0.00076	**0.01973**	**0.00066**
	100	0.02585	0.00106	0.02444	0.00088	0.02272	0.00080	**0.02105**	**0.00075**
	120	0.02679	0.00119	0.02381	0.00083	0.02302	0.00081	**0.02137**	**0.00076**

B. Glaxo Wellcome

Average Lag(n)		1		10		20		60	
		MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE
	1	0.11477	0.01977	0.05636	0.00416	0.05254	0.00337	0.05219	0.00294
	5	0.11403	0.01963	0.05765	0.00430	0.05315	0.00343	0.05203	0.00296
	10	0.11496	0.02000	0.05802	0.00437	0.05343	0.00339	0.05201	0.00299
Return	20	0.11428	0.01992	0.05767	0.00423	0.05312	0.00336	0.05176	0.00299
Squared	40	0.11584	0.02030	0.05925	0.00455	0.05431	0.00365	0.05230	0.00303
	60	0.11577	0.02034	0.06075	0.00470	0.05569	0.00373	0.05349	0.00310
	80	0.11616	0.02023	0.05895	0.00439	0.05530	0.00356	0.05517	0.00325
	100	0.11726	0.02022	0.06153	0.00476	0.05790	0.00392	0.05667	0.00351
	120	0.11764	0.02078	0.06361	0.00518	0.05987	0.00421	0.05725	0.00365
	1	0.10658	0.01832	0.04379	0.00267	0.03663	0.00178	0.03221	0.00125
	5	0.10586	0.01818	0.04408	0.00281	0.03728	0.00184	0.03211	0.00128
Return	10	0.10659	0.01855	0.04491	0.00288	0.03741	0.00181	0.03204	0.00130
Squared	20	0.10572	0.01848	0.04455	0.00275	0.03734	0.00179	0.03207	0.00132
Increased	40	0.10702	0.01884	0.04682	0.00305	0.03979	0.00205	0.03270	0.00134
by 0.02	60	0.10700	0.01884	0.04825	0.00316	0.04062	0.00208	0.03355	0.00136
	80	0.10719	0.01866	0.04484	0.00278	0.03809	0.00185	0.03532	0.00144
	100	0.10833	0.01859	0.04753	0.00308	0.04169	0.00215	0.03679	0.00164
	120	0.10857	0.01912	0.04994	0.00349	0.04356	0.00242	0.03802	0.00176
	1	0.09891	0.01762	0.03321	0.00192	0.02431	0.00090	0.01301	0.00024
	5	0.09836	0.01750	0.03337	0.00208	0.02415	0.00097	0.01374	0.00028
Return	10	0.09888	0.01786	0.03334	0.00215	0.02350	0.00094	0.01402	0.00031
Squared	20	0.09823	0.01781	0.03342	0.00203	0.02372	0.00093	0.01494	0.00034
Increased	40	0.09853	0.01815	0.03648	0.00230	0.02685	0.00117	0.01554	0.00034

by 0.0465*	60	0.09917	0.01808	0.03706	0.00234	0.02648	0.00113	0.01453	0.00029
	80	0.09861	0.01780	0.03247	0.00187	0.02238	0.00081	0.01359	0.00028
	100	0.09962	0.01766	0.03495	0.00210	0.02546	0.00103	0.01620	0.00040
	120	0.09962	0.01816	0.03707	0.00248	0.02749	0.00127	0.01704	0.00050
	1	0.09641	0.01781	0.03276	0.00208	0.02303	0.00099	0.01326	0.00027
	5	0.09607	0.01769	0.03365	0.00224	0.02330	0.00106	0.01394	0.00032
Return	10	0.09670	0.01805	0.03346	0.00231	0.02296	0.00104	0.01428	0.00034
Squared	20	0.09632	0.01801	0.03295	0.00220	0.02270	0.00103	0.01525	0.00038
Increased	40	0.09661	0.01833	0.03590	0.00246	0.02610	0.00126	0.01451	0.00036
by 0.06	60	0.09694	0.01823	0.03566	0.00247	0.02506	0.00119	0.01302	0.00029
	80	0.09590	0.01791	0.03132	0.00195	0.02114	0.00083	**0.01183**	**0.00023**
	100	0.09706	0.01773	0.03273	0.00214	0.02264	0.00100	**0.01423**	**0.00031**
	120	0.09673	0.01821	0.03454	0.00250	0.02476	0.00123	**0.01400**	**0.00039**
	1	**0.00509**	**0.00005**	0.00650	0.00008	0.00749	0.00011	0.01027	0.00020
	5	**0.00764**	**0.00012**	0.00844	0.00014	0.00889	0.00015	0.01091	0.00023
	10	**0.01007**	**0.00020**	0.01024	0.00019	**0.01010**	**0.00019**	0.01175	0.00025
Implied	20	0.01240	0.00028	0.01176	0.00026	**0.01145**	**0.00025**	0.01280	0.00030
Volatility	40	0.01332	0.00033	0.01243	0.00029	**0.01231**	**0.00027**	0.01348	0.00033
	60	0.01358	0.00030	0.01272	0.00026	**0.01211**	**0.00024**	0.01413	0.00034
	80	0.01271	0.00027	0.01209	0.00024	0.01249	0.00026	0.01524	0.00040
	100	0.01466	0.00037	0.01413	0.00035	**0.01401**	**0.00034**	0.01518	0.00041
	120	0.01573	0.00041	0.01526	0.00040	**0.01493**	**0.00036**	0.01565	0.00039

Notes: The results are based on 240 out-of-sample moving average forecasts, see subsection 3.3 for the moving average forecasts for implied volatility using IV and RS with the *n* most recent observations. The forecasting procedure is described in subsection 4.1. Bold numbers represent the smallest MAFE and the smallest MSFE for given forecasting horizons. In the case of a tie or a non-ranking, both are recorded in bold.

4.5 Comparison of Forecasting Performance of the Models

In this subsection, the results of the forecasting performance for all methods described in the section 3 are compared: GARCH(1,1)-RS, log-ARFIMA(0,*d*,1)-IV, detrended log-ARFIMA(0,*d*,1)-IV, scaled truncated log-ARFIMA(0,*d*,1)-IV, and the moving average method for the RS and IV. Table 2 shows MAFE and MSFE of six methods. Bold numbers report the smallest MAFE or the smallest MSFE for the given forecast horizons. As shown in subsections 4.3 and 4.4, the GARCH(1,1)-RS model and FIV_t^{60,RS^*} are not preferred to the log-ARFIMA(0,*d*,1)-IV model and $FIV_t^{20,IV}$. Thus, from now on, the following four models are considered: log-ARFIMA(0,*d*,1)-IV in sub-section 3.2.2, scaled truncated log-ARFIMA(0,*d*,1)-IV in sub-section 3.2.3, detrended log-ARFIMA(0,*d*,1)-IV in sub-section 3.2.4, and the moving average method for the IV in sub-section 3.3.

For short horizons, the long memory volatility models are preferred to $FIV_t^{20,IV}$. In this case, $FIV_t^{1,IV}$ will give smaller forecast errors than $FIV_t^{20,IV}$ (see table 5). The forecasting performances of $FIV_t^{1,IV}$ and the long memory volatility models are indistinguishable in short horizons. For long horizons, we may not differentiate the forecasting power of $FIV_t^{20,IV}$ from that of the detrended and scaled truncated log-ARFIMA(0,*d*,1)-IV models. Therefore, $FIV_t^{1,IV}$ and $FIV_t^{20,IV}$ can be used for the forecast of short and long horizons, respectively.

The forecasting performance of the detrended log-ARFIMA(0,*d*,1) model is reported in the third column of table 2. The detrended forecasts have less MAFE and MSFE than those of the log-ARFIMA(0,*d*,1)-IV model. Figures 2A and 3A suggest that the systematic forecast bias in the log-ARFIMA(0,*d*,1)-IV model can be reduced by this simple detrend method.

Despite the increase in FSTD in long forecasting horizons, table 2 and figures 2A to 3B suggest that the detrended method is not worse than the log-ARFIMA(0,d,1)-IV model, and performs well in long horizons.

The forecasting performance of the scaled truncated log-ARFIMA(0,d,1)-IV model is reported in the fourth column of table 2. The scaled truncated log-ARFIMA(0,d,1)-IV model performs well over all forecasting horizons. Figures 2A and 3A show that the scaled truncated log-ARFIMA(0,d,1)-IV model reduces the systematic forecast bias found in the log-ARFIMA(0,d,1)-IV model to a trivial level. In addition, the scaled truncated log-ARFIMA(0,d,1)-IV model reduces FSTD in long horizons; see figures 2B and 3B. Therefore, by reducing the systematic forecast bias and standard deviation, the scaled truncated log-ARFIMA(0,d,1)-IV model outperforms the log-ARFIMA(0,d,1)-IV in long forecasting horizons, while it holds the same forecasting power in short horizons as the log-ARFIMA(0,d,1)-IV. We suggest that the scaled truncated log-ARFIMA(0,d,1) model is preferred to the log-ARFIMA(0,d,1)-IV.

To make sure that our results are not dependent on the stock chosen or the time period, we selected 7 other stocks and FTSE100 index and three separate time periods. Although we only report 2 stocks for the period, the other results, available on request from the authors, are broadly similar and do not change our qualitative evaluations. However, we find that for some companies such as BTR, British Telecommunication, General Electric, and FTSE100 European call options, the log-ARFIMA(0,d,1)-IV model outperforms the scaled truncated log-ARFIMA(0,d,1)-IV for the forecast of implied volatility in long forecasting horizons. These implied volatilities have a common character that they have increasing trends during the forecasting period. In this case, the systematic forecasting bias in the log-ARFIMA(0,d,1)-IV model gives better forecasts. However, when an increasing trend in implied volatility is not anticipated, the scaled truncated log-ARFIMA(0,d,1)-IV performs well.

5. Conclusion

One of the referees raised an interesting philosophical issue as to whether RS is any use in forecasting IV. He notes that

> "If anything other than forecasted volatility affects the market price (e.g., risk aversion; expectations about the direction of future returns; supply and demand imbalances in the options market from liquidity trading and noise trading; effects of taxes, margin requirement, short sale constraints, and other "market imperfections"; etc., etc.), then implied volatility will impound potentially measurable explanatory factors that are not coming from expected volatility. Evidence of this comes from the fact that the author(s) find that implied volatility is systematically higher than historical volatility in the data. This means that if we really want to predict implied volatility, we are tying at least one hand behind our back by considering only models of returns volatility."

Our results rather support this view. We have used both RS data to forecast IV (GARCH(1,1)-RS model) and IV to forecast IV (log-ARFIMA(0,d,1)-IV) model). A referee has also pointed out that the GARCH(1,1)-RS model forecasts actual volatility, whilst the log-ARFIMA(0,d,1)-IV) model forecasts some temporal average of volatility over the remaining lifetime of the option. Thus, using GARCH models to forecast IV guarantees that the GARCH models will not do well. We acknowledge this point, however, to construct IV from a temporal average of the GARCH forecasts is not straightforward and we leave this for future research.

We can summarize our suggestion as follows. Firstly, for the forecast of implied volatility, IV rather than RS should be used. Secondly, log-ARFIMA(0,*d*,1)-IV is preferred to GARCH(1,1)-RS. Besides the forecasting performance results reported above, the log-ARFIMA(0,*d*,1) model does not need non-negativity constraints and estimates are easily obtained. Thirdly, the moving average method is not inferior to the more sophisticated methods such as GARCH and log-ARFIMA models for the forecast of long horizons. In addition, the estimate of d which is greater than 0.5 for our long memory models means that our models are actually random walks with some short memory correlation. Such a structure will favour short term forecasts, not long term forecasts. Finally, we also address the important issue of scaled truncation in ARFIMA(*k*,*d*,*l*) models and suggest a procedure that eliminates bias-induced trending in the forecasts whilst preserving the essential pattern of hyperbolic decay if it is present in the process. Our final recommendation for the forecast of implied volatility is scaled truncated ARFIMA(*k*,*d*,*l*) models for both short and long horizons.

Our evidence shows that the long memory in volatility may be eliminated by differencing, $ln(x_t^2) - ln(x_{t-1}^2)$. In this case, the growth rate in implied variance is covariance stationary with autocorrelation that decays exponentially. This means that whilst there is evidence of integration in IV models, there is no compelling evidence of long memory effects.

Finally, should financial houses use these models? It is by no means clear from our evidence. Over short horizons, these “fractional models” do well. However, over longer horizons, they only slightly outperform simple moving average forecasts. When costs are taken into account, simple methods may triumph.

Appendix

Appendix 1 <Mathematical Details of LM Models and Their Properties>

A simple ARIMA(0,1,0) model is defined as

(A1-1) $(1 - L)\, x_t = \varepsilon_t$

where ε_t is an independent identically distributed random variable. The equation (A1-1) means that the first difference of x_t is a discrete time white noise process. The idea of fractional integration permits the degree of difference to take any real value rather than integral values.

More formally, a fractionally integrated process is defined to be a discrete time stochastic process which is represented as

(A1-2) $\nabla^d x_t = (1 - L)^d x_t = \varepsilon_t$

The fractional difference operator ∇^d is defined by the binomial series expansion:

(A1-3)
$$\nabla^d = (1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} L^j$$

where $\Gamma(.)$ is the gamma function. Let $\gamma_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$. Then, via Stirlings approximation, it can be shown that $\gamma_j \approx \frac{j^{-d-1}}{\Gamma(-d)}$ as $j \to \infty$.

The autocovariance, autocorrelation and spectral density functions of the fractionally integrated process are[10]

[10] See Granger and Joyeux (1980) and Hosking (1981) for proof.

(A1-4)
$$\gamma_{FI}(s,d) = \frac{\Gamma(1-2d)\Gamma(s+d)}{\Gamma(d)\Gamma(1-d)\Gamma(s+1-d)}\sigma^2$$
$$\approx \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\sigma^2 s^{2d-1}, \qquad \text{as } s \to \infty,$$

(A1-5)
$$\rho_{FI}(s,d) = \frac{\Gamma(1-d)\Gamma(s+d)}{\Gamma(d)\Gamma(s+1-d)}$$
$$\approx \frac{\Gamma(1-d)}{\Gamma(d)} s^{2d-1}, \qquad \text{as } s \to \infty,$$

(A1-6)
$$S_{FI}(\omega,d) = \frac{\sigma^2}{2\pi}(2\sin(\frac{\omega}{2}))^{-2d}$$
$$\approx \frac{\sigma^2}{2\pi}\omega^{-2d}, \quad \text{as } \omega \to 0.$$

We can see that the autocorrelations of the fractionally integrated series decline at a slower rate than that of the ARMA model. The autocorrelation function (A1-5) decays at a hyperbolic rate, while that of the ARMA model decays exponentially.

Fractionally integrated processes show different characteristics depending on the parameter *d*. A fractionally integrated process is covariance stationary and invertible when $-0.5 < d < 0.5$, and it is a long memory process when *d* lies between 0 and 0.5. The fractional differencing parameter *d* is defined by the behaviour of the series up to infinite cycles. As *d* goes to 0.5, the decay rate of the impact of a unit innovation becomes slower. Hence, the fractional differencing parameter *d* decides the decay of the system's response to the innovation. Sowell (1990) shows that, while the variance of the partial sums of variables grows linearly with number of observations when $d = 0$, it grows faster than a linear rate when $0<d<0.5$. On the other hand, when $-0.5<d<0$, the process has short memory since each shock is negatively correlated with the others, thus making the variance of the partial sums of variables less than the variance of the individual shock.

Certain restrictions on the long memory parameter *d* are necessary for the process x_t to be stationary and invertible. The covariance stationarity condition needs the squared coefficients of the infinite order moving average representation to be summable. The moving average representation of equation (A1-2) is

(A1-7)
$$x_t = (1-L)^{-d}\varepsilon_t$$
$$= \sum_{j=0}^{\infty}\frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}\varepsilon_{t-j}$$

The variance of x_t can be represented as

(A1-8)
$$Var(x_t) = \sigma_\varepsilon^2[1+\sum_{j=1}^{\infty}(\frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)})^2]$$
$$\approx \sigma_\varepsilon^2[1+\Gamma(d)^{-2}\sum_{j=1}^{\infty}j^{2d-2}]$$

where σ_ε^2 is the variance of ε_t. Therefore, for the variance of x_t to exist, we need $2d-2<-1$ from the theory of infinite series. The long memory parameter which satisfies this condition is $d<0.5$. Thus, when $d<0.5$, x_t is a (weakly) stationary process. On the other hand, to obtain a convergent autoregressive representation of equation (A1-2), we can replace *d* in equation (A1-8) with $-d$. In this case, the invertibility condition is $-0.5<d$ for x_t.

Table A1-1 summarizes the properties of the long memory process for various *d* in the frequency domain context. Values of *d* outside the range $-0.5<d<0.5$ can be understood by differencing the series and examining the properties of the differenced process.

Table A1-1 Properties of the Fractionally Integrated Process in the Frequency Domain

d	S	I	Properties
d=-0.5	Yes	No	$s(\omega)$~0 as $\omega\rightarrow 0$
-0.5<d<0	Yes	Yes	short memory with negative correlation and high spectral density at high frequencies. $s(\omega)$~0 as $\omega\rightarrow 0$
d=0	Yes	Yes	white noise with zero correlation and constant spectral density. $s(\omega)=\sigma^2/2\pi$
0<d<0.5	Yes	Yes	long memory with positive correlation and high spectral density at low frequencies. $s(\omega)\sim\infty$ as $\omega\rightarrow 0$
d=0.5	No	Yes	$s(\omega)\sim\infty$ as $\omega\rightarrow 0$

Note: S and I represent stationarity and invertibility, respectively. $s(\omega)$ represents the spectral density function of the discrete time long memory process, see equation (A1-6).

Table A1-2 reports some examples of long memory coefficients at various lags. The key property of a long memory process is that its coefficients decay at a hyperbolical rate rather than the exponential rate of short memory process such as ARMA models. Therefore, the long memory process is a sensible process to describe high persistence in time series such as volatility.

Table A1-2 Comparison of Coefficients on Moving Average Representation between Long and Short Memory Processes

Lags	Fractionally Integrated Process				AR(1) Process			
	d=0.2	d=0.4	d=0.6*	d=0.8*	ϕ=0.2	ϕ=0.4	ϕ=0.6	ϕ=0.8
1	0.2000	0.4000	0.6000	0.8000	0.2000	0.4000	0.6000	0.8000
2	0.1200	0.2800	0.4800	0.7200	0.0400	0.1600	0.3600	0.6400
3	0.0880	0.2240	0.4160	0.6720	0.0080	0.0640	0.2160	0.5120
4	0.0704	0.1904	0.3744	0.6384	0.0016	0.0256	0.1296	0.4096
5	0.0591	0.1676	0.3444	0.6129	0.0003	0.0102	0.0778	0.3277
7	0.0454	0.1379	0.3031	0.5755	0.0000	0.0016	0.0280	0.2097
9	0.0372	0.1190	0.2752	0.5487	0.0000	0.0003	0.0101	0.1342
10	0.0342	0.1119	0.2642	0.5377	0.0000	0.0001	0.0060	0.1074
15	0.0248	0.0881	0.2255	0.4971	0.0000	0.0000	0.0005	0.0352
20	0.0197	0.0743	0.2014	0.4699	0.0000	0.0000	0.0000	0.0115
30	0.0143	0.0583	0.1716	0.4339	0.0000	0.0000	0.0000	0.0012
40	0.0114	0.0491	0.1531	0.4099	0.0000	0.0000	0.0000	0.0001
50	0.0095	0.0430	0.1401	0.3922	0.0000	0.0000	0.0000	0.0000
60	0.0082	0.0386	0.1303	0.3782	0.0000	0.0000	0.0000	0.0000
80	0.0065	0.0325	0.1162	0.3572	0.0000	0.0000	0.0000	0.0000
100	0.0055	0.0284	0.1063	0.3417	0.0000	0.0000	0.0000	0.0000
120	0.0047	0.0255	0.0988	0.3295	0.0000	0.0000	0.0000	0.0000
140	0.0042	0.0232	0.0929	0.3195	0.0000	0.0000	0.0000	0.0000
160	0.0038	0.0214	0.0881	0.3111	0.0000	0.0000	0.0000	0.0000
180	0.0034	0.0200	0.0841	0.3039	0.0000	0.0000	0.0000	0.0000
200	0.0031	0.0188	0.0806	0.2976	0.0000	0.0000	0.0000	0.0000
300	0.0023	0.0147	0.0686	0.2744	0.0000	0.0000	0.0000	0.0000
400	0.0018	0.0124	0.0611	0.2591	0.0000	0.0000	0.0000	0.0000
499	0.0015	0.0108	0.0559	0.2479	0.0000	0.0000	0.0000	0.0000

Notes: * means that the process is not stationary. The coefficients on the moving average representation of discrete time long memory processes are calculated using the following equation.

$$x_t = (1-L)^{-d}\varepsilon_t$$

$= \Sigma\, \gamma_j\, \varepsilon_{t-j}$

where $\gamma_j = \Gamma(j+d)/(\Gamma(j+1)\Gamma(d))$ and $\Gamma(.)$ is the Gamma Function. The coefficients on the moving average representation of AR processes are calculated using the following equation.

$$x_t = (1-\phi L)^{-1} e_t$$
$$= \Sigma\, \phi^{j}\, \varepsilon_{t-j}$$

Appendix 2 <The Definition and Properties of the FIGARCH and Log-ARFIMA Models>

Baillie, Bollerslev, and Mikkelsen (1996) introduce the concept of the fractional integration to GARCH models to make the following FIGARCH(p,d,q) model:

(A2-1) $\quad (1-\Phi(L))(1-L)^d y_t^2 = w + (1-\Theta(L))v_t$

where $0 \le d \le 1$, $v_t = y_t^2 - h_t$, $\Phi(L) = \phi_1 L + ,\dots, + \phi_q L^q$ and $\Theta(L) = \theta_1 L + ,\dots, + \theta_p L^p$. The conditional variance of the above FIGARCH model is expressed as

(A2-2) $\quad h_t = w + \Theta(L)h_t + [1-\Theta(L)-(1-\Phi(L))\,(1-L)^d]\; y_t^2$

In the FIGARCH model, the long memory parameter, d, is defined to have a value, $0 \le d \le 1$, while in the ordinary long memory return process, d is defined as $-0.5 < d < 0.5$ to be covariance stationary and invertible. In the FIGARCH model, d must not be less than zero because of the non-negativity conditions imposed on the conditional variance equation.

Note that there is a difference between the definition of the stationarity in the long memory return process and the long memory volatility process. In the long memory return process as shown in Appendix 1, the covariance stationary condition needs the summability of the squared moving average coefficients. However, in the FIGARCH model, the stationary condition depends on the summability of the moving average coefficients[11]. That is, stationarity in the FIGARCH model is defined as having an infinite moving average representation in L^1 space rather than L^2 space. The stationary condition in L^1 space is satisfied only when $d<0$. Therefore, when $0 \le d \le 1$, FIGARCH models are not covariance stationary.

Baillie, Bollerslev, and Mikkelsen (1996) suggest that FIGARCH models with $0 \le d \le 1$ are strictly stationary and ergodic by applying Bougerol and Picard (1992): IGARCH models are strictly stationary and ergodic. As explained in Baillie, Bollerslev, and Mikkelsen (1996), equation (A2-2) is equivalent to $h_t = (1-\Theta(1))^{-1} w + y_t^2$ at L=1. Therefore, w>0 in FIGARCH models can be interpreted in the same way as in IGARCH models, and the unconditional distribution of y_t^2 has infinite mean. This is a property of the long memory volatility process: every fractionally integrated volatility process with a drift does not have an unconditional distribution with finite mean[12]. This seems to be a major drawback as it says that, unconditionally, the expected value of implied volatility is infinite.

When equation (3-5) is combined together with conventional ARMA models, we can obtain ARFIMA(k,d,l) models. The model used in this study is a log-ARFIMA model which is represented as

(A2-3) $\quad (1-\Phi(L))(1-L)^d \ln(x_t^2) = \mu + (1+\Theta(L))\psi_t \qquad 0 \le d \le 1$

where $\Phi(L) = \phi_1 L + \phi_2 L^2 + ,\dots + \phi_k L^k$, and $\Theta(L) = \theta_1 L + \theta_2 L^2 + ,\dots + \theta_l L^l$, and ψ_t is a white noise

[11] See Baillie, Bollerslev, and Mikkelsen (1996) for further discussion.
[12] The following log-ARFIMA model has the same property.

zero mean process ($\psi_t = ln(x_t^2) - E_{t-1}(ln(x_t^2))$). The conditional log-variance of the log-ARFIMA model which follows from (A2-3) is

(A2-4) $$H_t = \mu + \Theta(L)\psi_t + (1-(1-\Phi(L))(1-L)^d)ln(x_t^2)$$
$$= \mu - \Theta(L)H_t + (1+\Theta(L)-(1-\Phi(L))(1-L)^d)ln(x_t^2)$$

where $H_t = E_{t-1}(ln(x_t^2))$. The log-ARFIMA model is defined as an ARFIMA model for log-variance and does not need non-negativity constraints. The above relationship expresses the conditional log-variance (H_t) in terms of lagged values of $ln(x_t^2)$ and H_t.

Note that equation (A2-4) is equivalent to $H_t = (1+\Theta(1))^{-1}\mu + ln(x_t^2)$ at L=1. In log-ARFIMA models, therefore, $\mu \neq 0$ has the same interpretation as in FIGARCH models. That is, the unconditional distribution of $ln(x_t^2)$ has infinite mean.

Appendix 3 <The Correction Factor of the Log-ARFIMA Model>

Notice that in log-ARFIMA models, the f-step-ahead conditional variance can not be represented as an exponential form of the f-step-ahead conditional log-variance. On the basis of Jensen's Inequality, the forecast $h_{t+f}^* = exp(H_{t+f})$ obtained from equation (3-10) is different from the appropriate forecast h_{t+f}. More formally,

(A3-1) $$\begin{aligned} h_{t+f} &= E_t(exp(ln\, y_{t+f}^2)) \\ &> exp(E(ln\, y_{t+f}^2)) = exp(H_{t+f}) = h_{t+f}^* \end{aligned}$$

If we assume that ψ_t is normal and define ζ_i to be the i-th coefficient of the moving average representation of the log-ARFIMA model, the appropriate forecast for the log-ARFIMA model is

(A3-2) $$\begin{aligned} h_{t+f} &= E_t(exp(ln\, y_{t+f}^2)) \\ &= E_t(exp(H_{t+f} + \sum_{i=0}^{f-1}\zeta_i\psi_{t+f-i}) \qquad (\zeta_0 = 1) \\ &= exp(H_{t+f})exp(\frac{1}{2}\sum_{i=0}^{f-1}\zeta_i^2\sigma_\psi^2) \end{aligned}$$

since for a normally distributed variable a, $E(exp(a)) = exp(E(a) + \frac{1}{2}var(a))$ where $E(a)$ and $var(a)$ are the mean and variance of a. Note that the correction factor, $exp(\frac{1}{2}\sum_{i=0}^{f-1}\zeta_i^2\sigma_\psi^2)$, is always larger than 1. Therefore, exp(H_{t+f}) gives downward biased forecasts, and the bias is an increasing function of the forecasting horizon, σ_ψ^2, and ζ_i.

Appendix 4 <The Systematic Forecast Bias>

Consider the following simple log-ARFIMA(0,d,0) model with a drift, $(1-L)^d\, ln(x_t^2) = \mu + \psi_t$. The process can be represented as

(A4-1) $$ln(x_t^2) = \mu - \sum_{j=1}^{\infty}\gamma_j\, ln(x_{t-j}^2) + \psi_t$$

where $\gamma_j = \dfrac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$. Note that with infinite observations, $-\sum_{j=1}^{\infty}\gamma_j = 1$ and unconditional log-variance does not exist. When we use a truncation lag, *m*, the process is represented as

(A4-2) $ln(x_t^2) = \tilde{\mu} - \sum_{j=1}^{m} \gamma_j \, ln(x_{t-j}^2) + \psi_t$

where $\tilde{\mu} = \mu - \sum_{j=m+1}^{\infty} \gamma_j \, ln(x_{t-j}^2)$. The drift, $\tilde{\mu}$, varies with *m*, *d*, and the magnitude of the log-variances beyond the truncation lag. Treating $\tilde{\mu}$ as a constant, we can obtain the following unconditional log-variance:

(A4-3) $E(ln(x_t^2)) = \dfrac{\tilde{\mu}}{\sum_{j=0}^{m} \gamma_j}$

where $0 < \sum_{j=0}^{m} \gamma_j$. Therefore, when we use a truncation lag, an unconditional log-variance exists.

The unconditional log-variance is achieved with a hyperbolic rate rather than an exponential rate as in GARCH and ARMA processes. Let A_f be a parameter on the drift term ($\tilde{\mu}$) in the *f*-step-ahead conditional log-variance of the log-ARFIMA(0,*d*,0) model. Then, A_f evolves hyperbolically as *f* increases as follows:

(A4-4) $A_1 = 1$, and $A_f = 1 - \sum_{j=1}^{f-1} \gamma_j A_{f-j}$, $f \geq 2$.

Therefore, the forecasts from the log-ARFIMA(0,*d*,1) model approach an unconditional variance with a slow decay rate.

Table A4-1 reports the sum of the AR coefficients, $-\sum_{j=1}^{m} \gamma_j$, over various values of *d* when a truncation is used. When *d* is close to 1, the sum of the AR coefficients becomes one for a relatively small truncation lag. However, it is far from 1 when *d* is small and *m* is moderate. When *d*=0.1, for example, the sum of the AR coefficients is far less than 1 even with the truncation lag of 10000 and we may obtain a large significant $\tilde{\mu}$, where $\tilde{\mu}$ is defined in equation (A4-2). In this case, applying such long memory processes with finite valued interpretations needs to be done in such a way as to preserve as many of the salient features of the theoretical process as possible.

Table A4-1 Sum of AR Coefficients of Fractionally Integrated Processes

Truncation Lag \ *d*	0.1	0.2	0.3	0.5	0.7	0.9	0.99
50	0.3678	0.6078	0.7623	0.9204	0.9784	0.9969	0.9998
100	0.4098	0.6583	0.8067	0.9437	0.9867	0.9983	0.9999
300	0.4711	0.7256	0.8609	0.9674	0.9938	0.9994	1.0000
500	0.4974	0.7522	0.8806	0.9748	0.9957	0.9996	1.0000
800	0.5204	0.7744	0.8963	0.9801	0.9969	0.9997	1.0000
1000	0.5310	0.7843	0.9030	0.9822	0.9973	0.9998	1.0000
1500	0.5496	0.8011	0.9141	0.9854	0.9980	0.9999	1.0000
2000	0.5624	0.8122	0.9212	0.9874	0.9984	0.9999	1.0000
2500	0.5721	0.8204	0.9263	0.9887	0.9986	0.9999	1.0000
3000	0.5798	0.8268	0.9302	0.9897	0.9988	0.9999	1.0000
5000	0.6007	0.8436	0.9402	0.9920	0.9991	1.0000	1.0000
7000	0.6139	0.8538	0.9459	0.9933	0.9993	1.0000	1.0000
10000	0.6275	0.8639	0.9514	0.9944	0.9995	1.0000	1.0000

Notes : A fractionally integrated process can be transformed into the following AR process:

$x_t = -\Sigma \gamma_j x_{t-j} + \varepsilon_t$

where $\gamma_j = \Gamma(j-d)/(\Gamma(j+1)\Gamma(-d))$ and $\Gamma(.)$ is the Gamma function. The numbers in the above table are sums of the AR coefficients for a given lag and *d*.

Appendix 5 <Scaled Truncated Long Memory Models>

The scaled truncated long memory model for a variable *z* is presented as

(A5-1) $(1-L)^{d_{ST}} z_t = \varepsilon_t$

The properties of the scaled truncated long memory process are clearly expressed in the following AR representation.

(A5-2) $z_t = \sum_{j=1}^{m} \gamma_j^* z_{t-j} + \varepsilon_t$

where $\gamma_j^* = \dfrac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(\text{-}d)} / \sum_{j=1}^{m} \dfrac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(\text{-}d)}$ and d is the original long memory parameter. Note that the sum of the scaled AR coefficients is always 1, $\sum_{j=1}^{m} \gamma_j^* = 1$, while $0 < -\sum_{j=1}^{m} \gamma_j < 1$ in equation (A4-2).

We shall now discuss the properties of the scaled truncated long memory process. The scaled truncated long memory process can be regarded as an AR(*m*) model with the sum of the AR coefficients constrained to be 1. However, in the scaled truncated long memory model, only one parameter, d_{ST}, is used for the long range dependence instead of *m* parameters as in the case of the AR(*m*) model. Furthermore, the decay rate retains the hyperbolic character associated with a long memory process. The invertibility conditions are the same as those of the ordinary fractionally integrated process in equation (3-5), since the AR coefficients in the scaled truncated long memory process are increased by a multiplication factor of $1/\sum_{j=1}^{m} \dfrac{\Gamma(j-d)}{\Gamma(j)\Gamma(\text{-}d)}$. Stationarity conditions will require checking if the roots of the appropriate polynomial lie outside the unit circle. There seems to be no results available on this question.

The scaled truncated fractionally integrated process does not result in the same degree of divergence between theory and practice as other forms of truncation imply for estimated models. In addition, it is worth noting that the long memory parameter of the scaled truncated fractionally integrated process is always less than the original long memory parameter for $0<d<1$. The gap between the two long memory parameters is smaller as *d* goes to 1 and vice versa. As the truncation lag increases, the long memory parameter of the scaled truncated fractionally integrated process will approach that of the ordinary fractionally integrated process. Therefore, with infinite samples and a truncation lag, the scaled truncated fractionally integrated process is equivalent to the ordinary fractionally integrated process.

Using the scaled truncated fractionally integrated process, we suggest the scaled truncated log-ARFIMA(*k*,*d*,*l*) model as follows:

(A5-3) $(1-\Phi(L))(1-L)^{d_{ST}} (ln(x_t^2) - \delta) = (1+\Theta(L))\psi_t$

where $\Phi(L) = \phi_1 L + \phi_2 L^2 + ... + \phi_k L^k$, and $\Theta(L) = \theta_1 L + \theta_2 L^2 + ... + \theta_l L^l$, and ψ_t is a white noise zero mean process ($\psi_t = ln(x_t^2) - E_{t-1}(ln(x_t^2))$). The conditional log-variance of the log-ARFIMA model which follows from (A5-3) is

(A5-4) $H_t = \delta + \Theta(L)\psi_t + (1-(1-\Phi(L))(1-L)^{d_{ST}})(ln(x_t^2) - \delta)$

where $H_t = E_{t-1}(ln(x_t^2))$. Therefore, the scaled truncated log-ARFIMA(0,*d*,1)-IV model is

(A5-5) $(1-L)^{d_{ST}} (ln(x_t^2) - \delta) = (1+\theta L)\psi_t$

and using the same method as in equation (3-10), the *f*-step-ahead conditional log-variance from the scaled truncated log-ARFIMA(0,d,1)-IV model is

(A5-6) $$H^{ST}_{t+f} = \delta - \sum_{j=1}^{f-1} \gamma^*_j (H^{ST}_{t+f-j} - \delta) - \sum_{j=f}^{m} \gamma^*_j (ln(x^2_{t+f-j}) - \delta) \qquad f \geq 2$$

where γ^*_j is defined in equation (A5-2) and with m is a truncation lag.

References

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996), Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 75(1), 3-30.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974), Estimation and Inference in Nonlinear Structure Models, *Annals of Economic and Social Measurement* 3/4, 653-665.

Black, F, and M. Scholes, 1972, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637-659.

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307-327.

Bollerslev, T., R. Y. Chou, and K. Kroner (1992), ARCH Modeling in Finance, *Journal of Ecomometrics* 52, 5-59.

Bollerslev, T. and J. M. Wooldridge (1992), Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances, *Econometric Reviews* 11(2), 143-172.

Bougerol, P., and N. Picard (1992), Stationarity of GARCH Processes and of Some Nonnegative Time Series, *Journal of Econometrics* 52, 115-128.

Day, E. T., and C. M. Lewis (1992), Stock Market Volatility and the Information Content of Stock Index Options, *Journal of Econometrics* 52, 267-287.

Diebold, F. X., and J. A. Lopez (1994), ARCH Models, mimeo, Department of Economics, University of Pennsylvania.

Ding, Z., C. Granger, and R. F. Engle (1992), A Long Memory Property of Stock Market Returns and a New Model, Discussion paper 92-21, Department of Economics, University of California, San Diego.

Engle, R. F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation, *Econometrica* 50(4), 987-1007.

Engle, R. F. and T. Bollerslev (1986), Modeling the Persistence of Conditional Variances, *Econometric Reviews* 5, 1-50.

Engle, R. F., C. Hong, A. Kane, and J. Noh (1993), Arbitrage Valuation of Variance Forecasts with Simulated Options, *Advances in Futures and Options Research* 6, 393-415.

Granger, C., and R. Joyeux (1980), An Introduction to Long Memory Time Series Models and Fractional Differencing, *Journal of Time Series Analysis* 1(1), 15-29.

Harvey, C. R., and R. E. Whaley (1991), S&P 100 Index Option Volatility, *Journal of Finance* 46(4), 1551-1561.

Harvey, C. R., and R. E. Whaley (1992), Market Volatility Prediction and the Efficiency of the S&P 100 Index Option Market, *Journal of Financial Economics* 31, 43-73.

Hosking, J. R. M. (1981), Fractional Differencing, *Biometrika* 68, 165-176.

Hwang, S., and S. Satchell (1997), Alternative Models for Implied and Return Volatilities, Outliers, and Further investigation of the Integration of Volatility, Unpublished Manuscript, Department of Applied Economics, University of Cambridge.

Lamoureux, C. G., and W. D. Lastrapes (1993), Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities, *Review of Financial Studies* 6(2), 293-326.

Mandelbrot, B., and J. Van Ness (1968), Fractional Brownian Motions, Fractional Noises and Applications, *SIAM Review* 10, 422-437

Noh, J., R. F. Engle, and A. Kane (1994), Forecasting Volatility and Option Prices of the S&P 500 Index, *Journal of Derivatives*, 17-30.

Psaradakis, Z., and M. Sola (1995), Modelling Long Memory in Stock Market Volatility: A Fractionally Integrated Generalised ARCH Approach, Discussion Paper in Financial Economics FE-2/95, Birkbeck College, University of London.

Sowell, F. (1990), The Fractional Unit Root Distribution, *Econometrica*, 58(2), 495-505.

Weiss, A. A. (1986), Asymptotic Theory for ARCH Models: Estimation and Testing, *Econometric Theory* 2, 107-131.